DIFFERENTIAL LABELING FOR QUANTITATIVE ANALYSIS OF COMPLEX PROTEIN MIXTURES

Related Applications

[0001] The present application claims priority to the U.S. Provisional Application Serial No. 60/305,232, filed July 13, 2001, by Haynes, et al., and entitled "DIFFERENTIAL LABELING FOR QUANTITATIVE ANALYSIS OF COMPLEX PROTEIN MIXTURES", and to U.S. Provisional Application Serial No. 60/264,576, filed January 26, 2001, by Haynes, et al., entitled "DIFFERENTIAL LABELING FOR QUANTITATIVE ANALYSIS OF COMPLEX PROTEIN MIXTURES", both of which are incorporated by reference herein in their entirety including any drawings.

Background of the Invention

[0002] Genomic technology has advanced to a point at which, in principle, it has become possible to determine complete genomic sequences and to quantitatively measure the mRNA levels for each gene expressed in a cell. For some species the complete genomic sequence has now been determined, and for one strain of the yeast *Saccharomyces cerevisiae*, the mRNA levels for each expressed gene have been precisely quantified under different growth conditions (Velculescu *et al.*, *Cell* 88:243-251 (1997)). Comparative cDNA array analysis and related technologies have been used to determine induced changes in gene expression at the mRNA level by concurrently monitoring the expression level of a large number of genes (in some cases all the genes) expressed by the investigated cell or tissue (Shalon *et al.*, *Genome Res* 6:639-645 (1996)). Furthermore, biological and computational techniques have been used to correlate specific function with gene sequences. The interpretation of the data obtained by these techniques in the context of the structure, control and mechanism of biological systems has been recognized as a considerable challenge. In particular, it has been extremely difficult to explain the mechanism of biological processes by genomic analysis alone.

[0003] Proteins are essential for the control and execution of virtually every biological process. The rate of synthesis and the half-life of proteins and thus their expression level are also controlled post-transcriptionally. Furthermore, the activity of proteins is frequently modulated by post-translational modifications, in particular protein phosphorylation, and dependent on the association of the protein with other molecules including DNA and proteins. Neither the level of expression nor the state of activity of proteins is therefore directly apparent from the gene sequence or even the expression level of the corresponding mRNA transcript. It is therefore essential that a complete description of a biological system include measurements that indicate the identity, quantity and the state of activity of the proteins which constitute the system. The large-scale (ultimately global) analysis of proteins expressed in a cell or tissue has been termed proteome analysis (Pennington *et al.*, *Trends Cell Bio* 7:168-173 (1997)).

[0004] At present no protein analytical technology approaches the throughput and level of automation of genomic technology. The most common implementation of proteome analysis is based on the separation of complex protein samples most commonly by two-dimensional gel electrophoresis (2DE) and the subsequent sequential identification of the separated protein species (Ducret *et al.*, *Prot Sci* 7:706-719 (1998); Garrels *et al.*, *Electrophoresis* 18:1347-1360 (1997); Link *et al.*, *Electrophoresis* 18:1314-1334 (1997); Shevchenko *et al.*, *Proc Natl Acad Sci USA* 93:14440-14445 (1996); Gygi *et al.*, *Electrophoresis* 20:310-319 (1999); Boucherie *et al.*, *Electrophoresis* 17:1683-1699 (1996)). This approach has been assisted by the development of powerful mass spectrometric techniques and the development of computer algorithms which correlate protein and peptide mass spectral data with sequence databases and thus rapidly identify proteins (Eng *et al.*, *J Am Soc Mass Spectrom* 5:976-980 (1994); Mann and Wilm, *Anal Chem* 66:4390-4399 (1994); Yates *et al.*, *Anal Chem* 67:1426-1436 (1995)). This technology (two-dimensional mass spectrometry) has reached a level of sensitivity which now permits the identification of essentially any protein which is detectable by conventional protein staining methods including silver staining (Figeys and Aebersold, *Electrophoresis* 19:885-892 (1998); Figeys *et al.*, *Nature Biotech* 14:1579-1583 (1996); Figeys *et al.*, *Anal Chem* 69:3153-3160 (1997); Shevchenko *et al.*, *Anal Chem* 68:850-858 (1996)). However, the sequential manner in

which samples are processed limits the sample throughput, the most sensitive methods have been difficult to automate and low abundance proteins, such as regulatory proteins, escape detection without prior enrichment, thus effectively limiting the dynamic range of the technique. In the 2DE/$(MS)^n$ method, proteins are quantified by densitometry of stained spots in the 2DE gels.

[0005] The development of methods and instrumentation for automated, data-dependent electrospray ionization (ESI) tandem mass spectrometry $(MS)^n$ in conjunction with microcapillary liquid chromatography (μLC) and database searching has significantly increased the sensitivity and speed of the identification of gel-separated proteins. As an alternative to the 2DE/$(MS)^n$ approach to proteome analysis, the direct analysis by tandem mass spectrometry of peptide mixtures generated by the digestion of complex protein mixtures has been proposed (Dongr'e *et al.*, *Trends Biotechnol* 15:418-425 (1997)). μLC-MS/MS has also been used successfully for the large-scale identification of individual proteins directly from mixtures without gel electrophoretic separation (Link *et al.*, *Nat Biotech*, 17:676-682 (1999); Opitek *et al.*, *Anal Chem* 69:1518-1524 (1997)). While these approaches accelerate protein identification, the quantities of the analyzed proteins cannot be easily determined, and these methods have not been shown to substantially alleviate the dynamic range problem also encountered by the 2DE/$(MS)^n$ approach. Therefore, low abundance proteins in complex samples are also difficult to analyze by the μLC/MS/MS method without their prior enrichment.

[0006] It is therefore apparent that current technologies, while suitable to identify a portion of the components of protein mixtures, are neither capable of measuring the quantity nor the state of activity of the protein in a mixture. Even improvements of the current approaches are unlikely to advance their performance sufficiently to make routine quantitative and functional proteome analysis a reality.

[0007] This invention provides methods and reagents that can be employed in proteome analysis which overcome the limitations inherent in traditional techniques The basic approach described can be employed for the quantitative analysis of protein expression in complex samples (such as cells, tissues, and fractions thereof), the detection and

quantitation of specific proteins in complex samples, and the quantitative measurement of specific enzymatic activities in complex samples.

[0008] In this regard, a multitude of analytical techniques are presently available for clinical and diagnostic assays which detect the presence, absence, deficiency or excess of a protein or protein function associable with a normal or disease state. While these techniques are quite sensitive, they do not necessarily provide chemical separation of products and may, as a result, be difficult to use for assaying several proteins or enzymes simultaneously in a single sample. Current methods may not distinguish among aberrant expression of different enzymes or their malfunctions which lead to a common set of clinical symptoms. The methods and reagents herein can be employed in clinical and diagnostic assays for simultaneously (multiplex) monitoring of multiple proteins and protein reactions.

[0009] Complex mixtures of proteins give rise to even more complex mixtures of peptides after proteolytic digestion. One way to reduce this complexity is to label a particular amino acid and then enrich for only those peptides containing the labeled amino acid. One good example of a selective peptide label is the use of iodoacetamido functional groups to specifically react with cysteine residues. Approximately 85-90% of all proteins contain at least one cysteine residue, which makes the labeling method applicable to almost all proteins present in a complex mixture. We have designed trifunctional synthetic peptide based reagents that can be used for reducing the complexity of peptide mixtures by labeling peptides with iodoacetamido groups and then selectively enriching only those peptides containing labeled cysteine residues.

Summary of the Invention

[0010] In the first aspect, the invention provides a compound of Formula I

(I) Immobilization Site-Cleavage Site-Link

where:

Immobilization Site is selected from the group consisting of an epitope tag, a linker to a solid surface, a metal chelating site, and a magnetic site, or a combination thereof;

Cleavage Site is selected from the group consisting of a protease cleavage site, a photocleavable linker, a restriction enzyme cleavage site, a chemical cleavage site, and a thermal cleavage site, or a combination thereof;

Link is selected from the group consisting of an amino acid reactive site and a mass variance site, or a combination thereof.

[0011] In another aspect, the invention provides a compound of Formula II or III:

(II) Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where:

A is an integer from 0 to 12;

X is selected from the group consisting of an amide bond of formula -C(O)-NR-, a carbonyl of formula -C(O)-, and an amino acid sequence comprising between 0 to 50 amino acids, where R is hydrogen or lower alkyl;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is selected from the group consisting of an amide bond of formula -$(CH_2)_B$-C(O)-NR-, an amide bond of formula -$(CH_2)_B$-NR-C(O)-, and an amino acid sequence comprising between 0 to 10 amino acids,

where R is hydrogen or lower alkyl, and

where B is an integer from 0 to 20;

alk is straight or branched chain of alkylene comprising between 0 and 20 carbon atoms;

Ph is a phenyl group optionally substituted with one or more electron withdrawing groups ortho or para to the -CH_2- group;

Link is selected from the group consisting of -$(CH_2)_C$-I, -$(CH_2)_D$-CH(-$(CH_2)_E CH_3$)-$(CH_2)_F$-X-I, Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide

where C, D, E, and F are each independently an integer from 0 to 20;

Epitope Tag Site is a sequence of amino acids,

where when A is two or more, the amino acid sequence of each Epitope Tag Site can be the same or different; and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme.

[0012] In another aspect, the invention provides for a method for simultaneously identifying and determining the levels of expression of cysteine-containing proteins in normal and perturbed cells, comprising:

a) preparing a first protein sample or a first peptide sample from the normal cells;

b) reacting the first protein sample or the first peptide sample with a reagent of Formula II or III:

(II) Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where:

A is an integer from 0 to 12;

X is selected from the group consisting of an amide bond of formula -C(O)-NR-, a carbonyl of formula -C(O)-, and an amino acid sequence comprising between 0 to 50 amino acids, where R is hydrogen or lower alkyl;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is selected from the group consisting of an amide bond of formula -$(CH_2)_B$-C(O)-NR-, an amide bond of formula -$(CH_2)_B$-NR-C(O)-, and an amino acid sequence comprising between 0 to 10 amino acids,

where R is hydrogen or lower alkyl, and

where B is an integer from 0 to 20;

alk is straight or branched chain of alkylene comprising between 0 and 20 carbon atoms;

Ph is a phenyl group optionally substituted with one or more electron withdrawing groups ortho or para to the $-CH_2-$ group;

Link is selected from the group consisting of $-(CH_2)_C-I$, $-(CH_2)_D-CH(-(CH_2)_ECH_3)-(CH_2)_F-X-I$, Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide

where C, D, E, and F are each independently an integer from 0 to 20;

Epitope Tag Site is a sequence of amino acids,

where when A is two or more, the amino acid sequence of each Epitope Tag Site can be the same or different; and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme;

c) preparing a second protein sample or a second peptide sample from the perturbed cells;

d) reacting the second protein sample or the second peptide sample of step c) with a second reagent of Formula II or III:

(II) Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where:

A is an integer from 0 to 12;

X is selected from the group consisting of an amide bond of formula -C(O)-NR-, a carbonyl of formula -C(O)-, and an amino acid sequence comprising between 0 to 50 amino acids, where R is hydrogen or lower alkyl;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is selected from the group consisting of an amide bond of formula $-(CH_2)_B-C(O)-NR-$, an amide bond of formula $-(CH_2)_B-NR-C(O)-$, and an amino acid sequence comprising between 0 to 10 amino acids,

where R is hydrogen or lower alkyl, and

where B is an integer from 0 to 20;

alk is straight or branched chain of alkylene comprising between 0 and 20 carbon atoms;

Ph is a phenyl group optionally substituted with one or more electron withdrawing groups ortho or para to the $-CH_2-$ group;

Link is selected from the group consisting of $-(CH_2)_C-I$, $-(CH_2)_D-CH(-(CH_2)_ECH_3)-(CH_2)_F-X-I$, Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide

where C, D, E, and F are each independently an integer from 0 to 20;

Epitope Tag Site is a sequence of amino acids,

where when A is two or more, the amino acid sequence of each Epitope Tag Site can be the same or different; and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme,

such that the molecular weight of the first reagent and the molecular weight of the second reagent are different by an integer multiple of 14 atomic mass units;

e) combining the reacted the first and the second protein samples or the reacted the first and the second peptide sample from steps b) and d);

f) subjecting the combined protein samples or the combined peptide samples from step e) to proteolysis at a site on the protein samples or at a site on the peptide samples, the site being other than the Protease Cleavage Site;

g) subjecting the proteolyzed combined protein samples or the proteolyzed peptide samples from step f) to an affinity chromatography system comprising a second amino acid sequence attached to a solid, thereby forming bound proteins and non-bound proteins,

where the Epitope Tag Site of the reagent and the second amino acid sequence bind with high specificity to each other;

h) eluting the non-bound proteins from the affinity chromatography system;

i) subjecting the affinity chromatography system from step h) to a protease specific for the Protease Cleavage Site, thereby forming a cleaved protein mixture;

j) eluting the cleaved protein mixture from the affinity chromatography system of step i);

k) isolating the eluted protein mixture obtained from step j);

l) subjecting the eluted protein mixture from step k) to chromatographic separation, followed by mass analysis;

m) comparing the results of step l) to:

1) determining the ratio of amounts of compounds in the two samples, where the molecular weights thereof are separated by an integer multiple of 14 atomic mass units; and

2) comparing the results obtained for each compound to protein databases containing chromatographic and molecular weight correlations.

[0013] In another aspect, the invention provides for a method for simultaneously identifying and determining the levels of expression of cysteine-containing proteins in normal and perturbed cells, comprising:

a) preparing a first protein sample or a first peptide sample from the normal cells;

b) subjecting the first protein sample or the first peptide sample from step a) to proteolysis;

c) reacting the proteolyzed first protein sample or the proteolyzed first peptide sample with a reagent of Formula II or III:

(II) Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where:

A is an integer from 0 to 12;

X is selected from the group consisting of an amide bond of formula -C(O)-NR-, a carbonyl of formula -C(O)-, and an amino acid sequence comprising between 0 to 50 amino acids, where R is hydrogen or lower alkyl;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is selected from the group consisting of an amide bond of formula $-(CH_2)_B$-C(O)-NR-, an amide bond of formula $-(CH_2)_B$-NR-C(O)-, and an amino acid sequence comprising between 0 to 10 amino acids,

where R is hydrogen or lower alkyl, and

where B is an integer from 0 to 20;

alk is straight or branched chain of alkylene comprising between 0 and 20 carbon atoms;

Ph is a phenyl group optionally substituted with one or more electron withdrawing groups ortho or para to the $-CH_2-$ group;

Link is selected from the group consisting of $-(CH_2)_C$-I, $-(CH_2)_D$-CH(-$(CH_2)_E CH_3$)-$(CH_2)_F$-X-I, Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide

where C, D, E, and F are each independently an integer from 0 to 20;

Epitope Tag Site is a sequence of amino acids,

where when A is two or more, the amino acid sequence of each Epitope Tag Site can be the same or different; and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme;

d) preparing a second protein sample or a second peptide sample from the perturbed cells;

e) subjecting the second protein sample or the second peptide sample from step d) to proteolysis;

f) reacting the proteolyzed second protein sample or the proteolyzed second peptide sample of step e) with a second reagent of Formula II or III:

(II) Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where:

A is an integer from 0 to 12;

X is selected from the group consisting of an amide bond of formula -C(O)-NR-, a carbonyl of formula -C(O)-, and an amino acid sequence comprising between 0 to 50 amino acids, where R is hydrogen or lower alkyl;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is selected from the group consisting of an amide bond of formula $-(CH_2)_B$-C(O)-NR-, an amide bond of formula $-(CH_2)_B$-NR-C(O)-, and an amino acid sequence comprising between 0 to 10 amino acids,

where R is hydrogen or lower alkyl, and

where B is an integer from 0 to 20;

alk is straight or branched chain of alkylene comprising between 0 and 20 carbon atoms;

Ph is a phenyl group optionally substituted with one or more electron withdrawing groups ortho or para to the $-CH_2-$ group;

Link is selected from the group consisting of $-(CH_2)_C$-I, $-(CH_2)_D$-CH(-$(CH_2)_E CH_3$)-$(CH_2)_F$-X-I, Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide

where C, D, E, and F are each independently an integer from 0 to 20;

Epitope Tag Site is a sequence of amino acids,

where when A is two or more, the amino acid sequence of each Epitope Tag Site can be the same or different; and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme,

such that the molecular weight of the first reagent and the molecular weight of the second reagent are different by an integer multiple of 14 atomic mass units;

g) combining the reacted the first and the second protein samples or the reacted the first and the second peptide sample from steps c) and f);

h) subjecting the combined protein samples or the combined peptide samples from step e) to proteolysis at a site on the protein samples or at a site on the peptide samples, the site being other than the Protease Cleavage Site;

i) subjecting the proteolyzed combined protein samples or the proteolyzed peptide samples from step f) to an affinity chromatography system comprising a second amino acid sequence attached to a solid, thereby forming bound proteins and non-bound proteins,

where the Epitope Tag Site of the reagent and the second amino acid sequence bind with high specificity to each other;

j) eluting the non-bound proteins from the affinity chromatography system;

k) subjecting the affinity chromatography system from step j) to a protease specific for the Protease Cleavage Site, thereby forming a cleaved protein mixture;

l) eluting the cleaved protein mixture from the affinity chromatography system of step k);

m) isolating the eluted protein mixture obtained from step l);

n) subjecting the eluted protein mixture from step m) to chromatographic separation, followed by mass analysis;

o) comparing the results of step n) to:

1) determining the ratio of amounts of compounds in the two samples, where the molecular weights thereof are separated by an integer multiple of 14 atomic mass units; and

2) comparing the results obtained for each compound to protein databases containing chromatographic and molecular weight correlations.

[0014] Another aspect of the present invention relates to a method for proteomic analysis, comprising:

a) preparing a protein sample or a peptide sample from cells;

b) reacting the protein sample or the peptide sample with a reagent of the formula:

Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

where:

A is an integer from 1 to 12;

X is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or X is an amino acid sequence comprising between 0 to 50 amino acids;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Z is an amino acid sequence comprising between 0 to 10 amino acids;

Link is selected from the group consisting of Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide;

Epitope Tag Site is a sequence of amino acids, and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme;

c) subjecting the reacted proteins or peptides from step b) to proteolysis at a site on the protein samples or at a site on the peptide samples, the site being other than the Protease Cleavage Site;

d) subjecting the proteolyzed reacted proteins or the proteolyzed reacted peptides from step c) to an affinity chromatography system comprising a second amino acid sequence attached to a solid support, thereby forming bound proteins and non-bound proteins,

where the Epitope Tag Site of the reagent and the second amino acid sequence bind with high specificity to each other;

e) eluting the non-bound proteins from the affinity chromatography system;

f) subjecting the affinity chromatography system from step e) to a protease specific for the Protease Cleavage Site, thereby forming a cleaved protein mixture;

g) eluting the cleaved protein mixture from the affinity chromatography system of step f);

h) isolating the cleaved protein mixture obtained from step g);

i) subjecting the cleaved protein mixture from step h) to chromatographic separation, followed by mass analysis;

j) comparing the results of step i) to:

1) determine the ratio of amounts of compounds in the sample separated by a molecular weight of 14 atomic mass units; and

2) identify the various modified proteins by comparing the results obtained for each modified protein to protein databases containing chromatographic and molecular weight correlations.

[0015] Yet another aspect of the invention relates to a process for preparing a fusion protein of the formula:

Protein-Acyl-N-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-[Lys-δ-N-iodoacetamide]

comprising,

a) preparing a fusion protein sample from cells having the formula Protein-Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Lys-δ-NHCOCH$_2$;

b) reacting the protein sample with an iodoacetamide,

where:

A is an integer from 1 to 12;

X is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or X is an amino acid sequence comprising between 0 to 50 amino acids;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Z is an amino acid sequence comprising between 0 to 10 amino acids;

Epitope Tag Site is a sequence of amino acids, and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme.

[0016] In another aspect, the invention relates to a process for preparing a fusion protein of the formula:

Protein-Acyl-N-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-[Orn-δ-N-iodoacetamide]

comprising,

a) preparing a fusion protein sample from cells having the formula Protein-Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Orn-δ-$NHCOCH_2$;

b) reacting the protein sample with an iodoacetamide,

where:

A is an integer from 1 to 12;

X is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or X is an amino acid sequence comprising between 0 to 50 amino acids;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Z is an amino acid sequence comprising between 0 to 10 amino acids;

Epitope Tag Site is a sequence of amino acids, and

Protease Cleavage Site is a sequence of amino acids that is a highly specific cleavage site for a protease enzyme.

Brief Description of the Drawings

[0017] Figure 1 is a chart showing the FPLC spectrum from the purification the synthesized PEPTag.

[0018] Figure 2a is a printout showing the mass spectrum of the synthesized PEPTag.

[0019] Figure 2b is a printout showing the mass spectrum from MS/MS experiment to sequence PEPTag.

[0020] Figures 3a,b show printouts of the MALDI MS analysis of PEPTag captured BSA peptides. Figure 3a is a printout wherein peaks are cysteinyl tryptic peptides from tagged BSA, which are captured by HA matrix and cleaved off by TEV. Figure 3b is a printout showing a control analysis of untagged BSA. The main peak in this spectrum is from TEV protease.

[0021] Figures 4a,b show the μLC MS/MS analysis of PEPTag captured BSA peptides. Figure 4a is a printout showing the base peak ion current profiles of all peptides released by TEV protease. Figure 4b is a printout showing the reconstructed ion

chromatograms from A (m/z 956.0-957.0) of the eluted peptide, which is doubly charged ion (m/z=956.4).

[0022] Figures 5a,b show the MS and MS/MS spectra of the PEPTag modified peptide. Figure 5a is a printout showing the full-scan (600-1,500 m/z) mass spectrum at time 29.49 min of μLC-MS and μLC-MS/MS analysis. Figure 5b is a printout showing the tandem mass spectrum (250-1925 m/z) of the $(M+2H)^{2+}$ of the eluted peptide (m/z=957.25).

[0023] Figure 6 is a printout showing the MALDI mass spectrum of a pair of PEPTag labeled peptides of identical sequences. The m/z difference depends on the charge state. It is either 14 or 7 for charge state one or two.

[0024] Figures 7a-c show the μLC-MS/MS analysis of captured peptides labeled by differential PEPTags. Figure 7a is a printout showing base peak ion current profiles of all the peptides released by TEV protease from combined two protein mixtures. Figure 7b is a printout showing the reconstructed ion chromatograms (m/z 1034.0-1035.0) of a cysteinyl peptide labeled by PEPTag 1a. Figure 7c is a printout showing the reconstructed ion chromatograms (m/z 1027.0-1028.0) of the same cysteinyl peptide labeled by PEPTag 1b.

[0025] Figure 8 is a printout of the ESI mass spectrum of the pair of PEPTag labeled peptides of identical sequences. The m/z difference is 7 for doubly charged ions.

Detailed Description of the Preferred Embodiments

[0026] Embodiments of this invention provide analytical reagents and mass spectrometry-based methods using these reagents for the rapid and quantitative analysis of proteins or protein function in mixtures of proteins. The analytical method can be used for qualitative and particularly for quantitative analysis of global protein expression profiles in cells and tissues, *i.e.*, the quantitative analysis of proteomes. The method can also be employed to screen for and identify proteins whose expression level in cells, tissue or biological fluids is affected by a stimulus (*e.g.*, administration of a drug or contact with a potentially toxic material), by a change in environment (*e.g.*, nutrient level, temperature, passage of time) or by a change in condition or cell state (*e.g.*, disease state, malignancy, site-directed mutation, gene knockouts) of the cell, tissue or organism from which the sample originated. The proteins identified in such a screen can function as markers for the changed

state. For example, comparisons of protein expression profiles of normal and malignant cells can result in the identification of proteins whose presence or absence is characteristic and diagnostic of the malignancy.

[0027] In an exemplary embodiment, the methods herein can be employed to screen for changes in the expression or state of enzymatic activity of specific proteins. These changes may be induced by a variety of chemicals, including pharmaceutical agonists or antagonists, or potentially harmful or toxic materials. The knowledge of such changes may be useful for diagnosing enzyme-based diseases and for investigating complex regulatory networks in cells.

[0028] The methods herein can also be used to implement a variety of clinical and diagnostic analyses to detect the presence, absence, deficiency or excess of a given protein or protein function in a biological fluid (*e.g.*, blood), or in cells or tissue. The method is particularly useful in the analysis of complex mixtures of proteins, *i.e.*, those containing 5 or more distinct proteins or protein functions.

[0029] One method employs affinity-labeled protein reactive reagents that allow for the selective isolation of peptide fragments or the products of reaction with a given protein (*e.g.*, products of enzymatic reaction) from complex mixtures. The isolated peptide fragments or reaction products are characteristic of the presence of a protein or the presence of a protein function, *e.g.*, an enzymatic activity, respectively, in those mixtures. Isolated peptides or reaction products are characterized by mass spectrometric (MS) techniques. In particular, the sequence of isolated peptides can be determined using tandem MS $(MS)^n$ techniques, and by application of sequence database searching techniques, the protein from which the sequenced peptide originated can be identified.

I. Reagents of the Invention

[0030] Embodiments of the present invention provide trifunctional synthetic reagents that can be used for reducing the complexity of peptide mixtures by labeling peptides at a specific amino acid residue and then selectively enriching only those peptides containing the labeled amino acid. By preparing this reagent in two forms with detectably

different masses, this technique can be used to provide accurate relative quantification of peptide amounts using mass spectrometry.

[0031] The amino acids used in the reagents of the present invention may be the D isomer or the L isomer of the amino acid. Thus, the one-letter designation "A" or the three-letter designation "ala," for example, refers to both D-alanine and L-alanine. In addition, the amino acids used in the reagents of the present invention may be naturally occurring or synthetic. Thus, for example, the one-letter designation "A" or the three-letter designation "ala," refers to both the naturally occurring alanine, having the formula $^{+}H_3N$-$CH(CH_3)$-COO^{-}, or any chemically modified analog thereof.

[0032] In some embodiments of the invention, the peptide labeling moiety consists of a lysine residue modified with an iodoacetamide functional group on the ε-amino group of the side chain. The synthetic peptides contain two additional motifs: a peptide epitope tag for high affinity purification; and a highly specific protease site for releasing the affinity purified labeled peptides from the affinity matrix. In addition, these synthetic peptides can readily be prepared as isoforms of two different masses by the simple expedient of using an ornithine in place of lysine to introduce a 14 mass unit difference in the carboxyl terminal acid.

[0033] In other embodiments of the invention, the peptide labeling moiety consists of a molecule modified with an iodo-containing organic substituent, which may be an iodide on a primary carbon, an acid iodide, or an iodoacetamide functional group. In addition, the peptide labeling moiety comprises a substituted benzyl moiety, which undergoes heterolytic cleavage upon exposure to light of a certain wavelength. In addition, these molecules can readily be prepared as isoforms of two different masses by the simple expedient of using an alkylene chain that has additional methylene groups or is missing methylene groups to introduce an integer multiple of 14 mass unit difference in the carboxyl terminal acid.

[0034] Thus, in a first aspect, the invention provides a compound of Formula I

(I) Immobilization Site-Cleavage Site-Link

where:

Immobilization Site is selected from the group consisting of an epitope tag, a linker to a solid surface, a metal chelating site, a magnetic site, and a specific oligonucleotide sequence, or a combination thereof;

Cleavage Site is selected from the group consisting of a protease cleavage site, a photocleavable linker, a restriction enzyme cleavage site, a chemical cleavage site, and a thermal cleavage site, or a combination thereof;

Link is selected from the group consisting of an amino acid reactive site and a mass variance site, or a combination thereof.

[0035] At some point during their use, the compounds of the present invention are immobilized on, for example, a surface, such that they do not move when washed with a fluid. The surface on which the compounds are immobilized may be a solid surface. Examples, without limitation of solid surfaces include beads (glass, plastic or other material), plastic, glass, silicon chip, multi-well plates, and membranes (such as PVDF or nylon).

[0036] There are a number of ways by which the compounds of the invention may be immobilized. For instance, the solid surface may comprise an amino acid sequence. The Immobilization Site of the compounds of the present invention will then comprise another amino acid sequence which is the epitope tag of the amino acid sequence on the surface. An epitope tag binds exclusively to its target amino acid sequence.

[0037] In other embodiments, the solid surface may comprise a metal chelating column, comprising for example nickel atoms. The Immobilization Site of the compounds of the invention may then comprise, for example, amino acid residues, such as histidines, or other residues, such as ethylenediaminetetraacetate, that will chelate to the metal atom on the column. The solid surface can be an oligonucleotide and the Immobilization Site can be the complimentary oligonucleotide. Those skilled in the art and familiar with metal affinity chromatography will know which chelating groups are best used with which metals on the column to be used.

[0038] In other embodiments of the present invention, the solid surface may comprise magnetic residues. In this case, the Immobilization Site of the compounds of the present invention will also comprise magnetic residues that are designed to bind magnetically to the magnetic residues of the solid surface.

[0039] In certain other embodiments, the Immobilization Site is a direct link between the solid surface and the compounds of the present invention. The direct link may be an acyl group or other chemical moieties that are capable of reacting with the solid surface, in some cases reversibly, so that the compounds of the present invention are immobilized on the surface.

[0040] The Cleavage Site is a part of the compound of the present invention that is capable of breaking the molecule in two different parts: One part of the molecule remains immobilized on the solid surface, while the other part of the molecule can move away from the solid surface by a wash fluid.

[0041] In certain embodiments, the Cleavage Site may be an amino acid sequence, comprising at least one amino acid residue, which is a cleavage site for a protease.

[0042] In other embodiments, the Cleavage Site may be a photocleavable linker. A photocleavable linker is a residue that breaks in two parts, either heterolytically or homolytically, when exposed to light of a certain wavelength, whether visible, infrared, or ultraviolet.

[0043] Other embodiments of the invention include a Cleavage Site which comprises a polynucleotide residue, of at least two nucleotides in length, that can be cleaved with a restriction enzyme.

[0044] In certain other embodiments, the Cleavage Site is a site that can be chemically cleaved, for example, by addition of an acid or a base.

[0045] In other embodiments, the Cleavage Site may be cleaved thermally. This embodiment may include a Cleavage Site that comprises a polynucleotide reside that can hybridize to another polynucleotide residue connected to the Immobilization Site. Heating the compounds can then result in the hybridized polynucleotides to "melt" and separate, as a DNA double helix would.

[0046] The Link comprises a residue that can react with an amino acid. The Link may react with a side-chain of an amino acid, or with the N- or C-terminus of a polypeptide. Thus, the Link residue comprises a reactive group. The reactive group may be a moiety that can undergo nucleophilic substitution with a portion of the amino acid, or can form an amide

or an ester bond with the amino acid. However, in general, the invention contemplates any reactive group that can form a bond with any part of an amino acid.

[0047] Optionally, the Link comprises a portion that allows mass variance to be introduced into a series of molecules. Thus, for example, the Link residue comprises a alkylene group, which may be a methylene in one embodiment, an ethylene in another embodiment, and a propylene in yet another embodiment, thereby introducing a mass difference of a multiple of 14 mass units between the different embodiments. The mass variance portion of the Link residue may be a series of methylene residues, or a series of -NH- residues, or a series of amide bonds, -NH-C(O)-. Any other repeating unit may work for introducing mass variance. The mass variance may be a variance that is measurable under the conditions of the experiment. Thus, mass variances in the range of 1 to 1000 mass units, or in the range of about 1 to about 500 mass units, or in the range of about 1 to about 250 mass units, or in the range of about 1 to about 100, or in the range of about 1 to about 50, or in the range of about 1 to about 30, or in the range of about 1 to about 20, or in the range of about 3 to about 20, or in the range of about 4 to about 20 are contemplated. In general, the mass variance portion of the Link affects chromatographic properties of the compound of the invention consistently.

[0048] In another aspect, the invention provides a compound of Formula II or III:

(II) Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where:

A is an integer from 0 to 12;

X is selected from the group consisting of an amide bond of formula -C(O)-NR-, a carbonyl of formula -C(O)-, and an amino acid sequence comprising between 0 to 50 amino acids, where R is hydrogen or lower alkyl;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is selected from the group consisting of an amide bond of formula $-(CH_2)_B-C(O)-NR-$, an amide bond of formula $-(CH_2)_B-NR-C(O)-$, and an amino acid sequence comprising between 0 to 10 amino acids,

where R is hydrogen or lower alkyl, and

where B is an integer from 0 to 20;

alk is straight or branched chain of alkylene comprising between 0 and 20 carbon atoms;

Ph is a phenyl group optionally substituted with one or more electron withdrawing groups ortho or para to the $-CH_2-$ group;

Link is selected from the group consisting of $-(CH_2)_C-I$, $-(CH_2)_D-CH(-(CH_2)_ECH_3)-(CH_2)_F-X-I$, Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide

where C, D, E, and F are each independently an integer from 0 to 20;

Epitope Tag Site is a sequence of amino acids,

where when A is two or more, the amino acid sequence of each Epitope Tag Site can be the same or different; and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme.

[0049] By "Acyl" it is meant a chemical substituent of the formula R-C(O)-, where R is an organic group selected from the group consisting of straight chain, branched, or cyclic alkyl, aryl, and five-membered or six-membered heteroaryl, each being optionally substituted with one or more protected substituents, which are selected from the group consisting of hydroxyl (-OH), sulfhydryl (-SH), amino ($-NH_2$), nitro ($-NO_2$), carboxyl (-COOH), ester (-COOR), and carboxamido ($-CONH_2$). These substituents may be protected by any common organic protecting group as set forth in, for example, Greene & Wutts, Protective Groups in Organic Chemistry, 3rd Ed., John Wiley & Sons, New York, NY, 1999.

[0050] Electron withdrawing groups are well-known to those of skill in the art. These groups include, without limitation, -OH, -OR, $-NO_2$, $-N(CH_3)_3^+$, -CN, -COOH, -COOR, $-SO_3H$, -CHO, and -CRO. In general, these groups are the ones that increase the rate

of nucleophilic aromatic substitution when they are located at the ortho or para position with respect to the site of attack.

[0051] One of the functional groups of the compounds is the Epitope Tag Site. Suitable Epitope Tag Sites bind selectively either covalently or non-covalently and with high affinity to a capture reagent. The "capture reagent" is an amino acid sequence bound to solid support. The solid support, with the capture reagent attached thereto, are packed into a column, preferably a column for chromatography. The amino acid sequence of the capture reagent and the amino acid sequence of the Epitope Tag Site are designed to bind to each other with high selectivity and high affinity. The binding may be either covalently or non-covalently. Examples of non-covalent binding include ionic interactions, van der Waals interactions, and hydrophobic or hydrophilic interactions. The binding between the Epitope Tag Site and the capture reagent may be similar to the binding of an antibody to an epitope of a protein for which the antibody is specific.

[0052] The interaction or bond between the Epitope Tag Site and the capture agent preferably remains intact after extensive and multiple washings with a variety of solutions to remove non-specifically bound components. The Epitope Tag Site binds minimally or preferably not at all to components in the assay system, except the capture agent, and does not significantly bind to surfaces of reaction vessels. Any non-specific interaction of the Epitope Tag Site with other components or surfaces should be disrupted by multiple washes that leave Epitope Tag Site-capture agent interaction intact. Further, the interaction of Epitope Tag Site and the capture agent can be disrupted to release peptide, substrates or reaction products, for example, by addition of a displacing ligand or by changing the temperature or solvent conditions. Preferably, neither capture agent nor Epitope Tag Site react chemically with other components in the assay system and both groups should be chemically stable over the time period of an assay or experiment.

[0053] The Epitope Tag Site is preferably soluble in the sample liquid to be analyzed and the capture reagent should remain soluble in the sample liquid even though attached to an insoluble resin such as Agarose. In the case of the capture reagent, the term "soluble" means that the capture reagent is sufficiently hydrated or otherwise solvated such that it functions properly for binding to the Epitope Tag Site. The capture reagent or capture

reagent-containing conjugates should not be present in the sample to be analyzed, except when added to capture the Epitope Tag Site.

[0054] A displacement ligand is optionally used to displace the Epitope Tag Site from the capture reagent. Suitable displacement ligands are not typically present in samples unless added. The displacement ligand should be chemically and enzymatically stable in the sample to be analyzed and should not react with or bind to components (other than the capture reagent) in samples or bind non-specifically to reaction vessel walls. The displacement ligand preferably does not undergo peptide-like fragmentation during mass spectral analysis, and its presence in sample should not significantly suppress the ionization of tagged peptide, substrate or reaction product conjugates.

[0055] Another functional group of the compounds disclosed herein is the Protease Cleavage Site. This site is an amino acid sequence, which in some embodiments comprises between 1 and 15 amino acids, and in other embodiments comprises between 4 and 8 amino acids, while in certain other embodiments comprises at least four amino acids. In one embodiment, the Protease Cleavage Site is an amino acid sequence of formula ENLYFQG (SEQ ID NO: 1).

[0056] The Protease Cleavage Site is designed to be cleaved once it is exposed to a highly specific protease enzyme. In certain embodiments, the protease enzyme is selected from the group consisting of TEV protease, chymotrypsin, endoproteinase Arg-C, endoproteinase Asp-N, trypsin, *Staphylococcus aureus* protease, thermolysin, and pepsin. In other embodiments, the protease enzyme is TEV protease. Preferably, the Protease Cleavage Site is not cleaved by the enzyme for the initial proteolysis of the lysed cell sample, nor would the cleavage site be lysed by any contaminating proteases from the cell sample.

[0057] The third functional group of the compounds disclosed herein is the protein reactive group, designated as "Link" in the above formula. This group may selectively react with certain protein functional groups or may be a substrate of an enzyme of interest. Any selectively reactive protein reactive group should react with a functional group of interest that is present in at least a portion of the proteins in a sample. Reaction of Link with functional groups on the protein should occur under conditions that do not lead to substantial degradation of the compounds in the sample to be analyzed. Examples of

selectively reactive Links suitable for use in the affinity tagged reagents include those which react with sulfhydryl groups to tag proteins containing cysteine, those that react with amino groups, carboxylate groups, ester groups, phosphate reactive groups, and aldehyde and/or ketone reactive groups or, after fragmentation with CNBr, with homoserine lactone.

[0058] Thiol reactive groups include epoxides, α-haloacyl groups, nitriles, sulfonated alkyls or aryl thiols and maleimides. Amino reactive groups tag amino groups in proteins and include sulfonyl halides, isocyanates, isothiocyantes, active esters, including tetrafluorophenyl esters, and N-hydroxysuccinimidyl esters, acid halides, and acid anyhydrides. In addition, amino reactive groups include aldehydes or ketones in the presence or absence of $NaBH_4$ or $NaCNBH_3$.

[0059] Carboxylic acid reactive groups include amines or alcohols in the presence of a coupling agent such as dicyclohexylcarbodiimide, or 2,3,5,6-tetrafluorophenyl trifluoroacetate and in the presence or absence of a coupling catalyst such as 4-dimethylaminopyridine; and transition metal-diamine complexes including Cu(II)phenanthroline.

[0060] Ester reactive groups include amines which, for example, react with homoserine lactone.

[0061] Phosphate reactive groups include chelated metal where the metal is, for example Fe(III) or Ga(III), chelated to, for example, nitrilotriacetiac acid or iminodiacetic acid.

[0062] Aldehyde or ketone reactive groups include amine plus $NaBH_4$ or $NaCNBH_3$, or these reagents after first treating a carbohydrate with periodate to generate an aldehyde or ketone.

[0063] The Link group should be soluble in the sample liquid to be analyzed and it should be stable with respect to chemical reaction, *e.g.*, substantially chemically inert, with components of the sample as well as the Epitope Tag Site, Protease Cleavage Site, and the capture reagent groups. The Link group when bound to the molecule should not interfere with the specific interaction of the Epitope Tag Site with the capture reagent or interfere with the displacement of the Epitope Tag Site from the capture reagent by a displacing ligand or by a change in temperature or solvent. The Link group should bind minimally or preferably

not at all to other components in the system, to reaction vessel surfaces or to the capture reagent. Any non-specific interactions of the Link group should be broken after multiple washes which leave the Epitope Tag Site-capture reagent complex intact.

[0064] The Link group may be selected from a group of substituents that differ from one another by the presence or absence of one or more repeating units, such as methylene ($-CH_2-$) groups. Thus, groups that contain straight chain alkylene moieties within them are particularly well-suited for this purpose.

[0065] In certain embodiments, the invention contemplates using lysine, ornithine, or arginine, coupled with iodoacetamide, as the Link group. "Orn" is the three letter designation for "L-ornithine," which is (S)-(+)-2,5-diaminopentanoic acid, $H_2N(CH_2)_3CH(NH_2)COOH$. "Iodoacetamide" is an organic substituent group with the structure $I-CH_2-C(O)-NH-$. When an amino acid group of a compound is derivatized by the iodoacetamide group, the iodoacetamide group is chemically bound to the side-chain amino group of the amino acid moiety. Thus, the designation "ε" or "δ" following the amino acids in the above formula designate the position at which the amino acid is derivatized by the iodoacetamide group. For example, Lys-ε-iodoacetamide has the formula

$$ICH_2C(O)NH(CH_2)_4CH(NH_2)COOH$$

[0066] It is also understood within the context of the invention that the incorporation of the designation "ε" or "δ" is optional. Therefore, Lys-ε-iodoacetamide and Lys-iodoacetamide (K-iodoacetamide), Arg-δ-iodoacetamide and Arg-iodoacetamide (R-iodoacetamide), and Orn-δ-iodoacetamide and Orn-iodoacetamide refer to the same compound or moiety, respectively.

[0067] Specific embodiments provided herein include, but are in no way limited to, the following compounds:

Acyl-NH-AYPYDVPDYASENLYFQGK-iodoacetamide (SEQ ID NO: 2),

Acyl-NH-AYPYDVPDYASENLYFQGGK-iodoacetamide (SEQ ID NO: 3),

Acyl-NH-AYPYDVPDYASENLYFQGAK-iodoacetamide (SEQ ID NO: 4),

Acyl-NH-AYPYDVPDYASENLYFQG(GABA)K-iodoacetamide (SEQ ID NO: 5),

Acyl-NH-AYPYDVPDYASENLYFQGVK-iodoacetamide (SEQ ID NO: 6),

Acyl-NH-AYPYDVPDYASENLYFQGOrn-iodoacetamide (SEQ ID NO: 7),

Acyl-NH-AYPYDVPDYASENLYFQGGOrn-iodoacetamide (SEQ ID NO: 8),
Acyl-NH-AYPYDVPDYASENLYFQGAOrn-iodoacetamide (SEQ ID NO: 9),
Acyl-NH-AYPYDVPDYASENLYFQG(GABA)Orn-iodoacetamide (SEQ ID NO: 10),
Acyl-NH-AYPYDVPDYASENLYFQGVOrn-iodoacetamide (SEQ ID NO: 11),
Acyl-NH-AYPYDVPDYASENLYFQGR-iodoacetamide (SEQ ID NO: 12),
Acyl-NH-AYPYDVPDYASENLYFQGGR-iodoacetamide (SEQ ID NO: 13),
Acyl-NH-AYPYDVPDYASENLYFQGAR-iodoacetamide (SEQ ID NO: 14),
Acyl-NH-AYPYDVPDYASENLYFQG(GABA)R-iodoacetamide (SEQ ID NO: 15), and
Acyl-NH-AYPYDVPDYASENLYFQGVR-iodoacetamide (SEQ ID NO: 16).

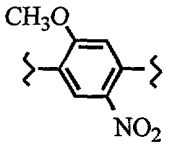
[0068] Other specific embodiments include:
Acyl-NH-CASENLYFQGK-$CH_2CH_2CH_2CH_2$-NH-C(O)-CH_2I,
Acyl-NH-CASENLYFQGOrn-$CH_2CH_2CH_2$-NH-C(O)-CH_2I,
Acyl-NH-CASENLYFQGPK-$CH_2CH_2CH_2CH_2$-NH-C(O)-CH_2I, and
Acyl-NH-CASENLYFQGPOrn-$CH_2CH_2CH_2CH_2$-NH-C(O)-CH_2I.

[0069] Other embodiments of the invention include compounds in which the Link moiety is a non-amino acid organic group. In these embodiments, the Link moiety is $-(CH_2)_C$-I or $-(CH_2)_D$-CH(-$(CH_2)_ECH_3$)-$(CH_2)_F$-X-I, where C, D, E, and F are each independently an integer from 0 to 20, and X is as defined herein. In some embodiments, the Link group is iodoacetamide. In other embodiments, the Link group is selected from the group consisting of $-CH(CH_2C(O)I)CH_2CH_3$, $-C(C(O)I)CH_2CH_2CH_3$, $-CH(CH_2I)CH_2CH_3$, $-CH_2CH(CH_2I)CH_2CH_2CH_3$.

[0070] In other embodiments, the invention relates to a compound of Formula III. In some embodiments, alk is a straight or branched chain of alkylene comprising between 0 and 20, between 0 and 15, between 0 and 10, between 0 and 5, or between 0 and 3 carbon atoms carbon atoms. In some embodiments alk is a straight chain of alkylene. alk may be selected from the group consisting of methylene, ethylene, propylene, n-butylene, and n-pentylene. In certain embodimets, alk is propylene.

[0071] In some embodiments Ph is a substituted phenyl group. It may be substituted with electron withdrawing groups. The substitutions may take place at positions

ortho or para to the methylene group to which Ph is connected. In certain embodiments, the substituents on Ph are methoxy or nitro. In some embodiments, Ph is the following:



[0072] The Ph groups is such that when the molecule is exposed to a light of certain wavelength, for example ultraviolet light, the bond between the CH_2 group and Z undergoes heterolytic cleavage. Therefore, the substituents on Ph are situated to stabilize the resulting benzylic free radical.

[0073] In embodiments, Z is an amino acid sequence comprising between 1 and 3 amino acids. In certain embodiments, Z is a single amino acid. It may be any of the natural or synthetic amino acids known in the art. In some embodiments, Z is selected from the group consisting of glycine, alanine, and valine. In certain other embodiments, Z may be a synthetic amino acid, where the amino group in a position other than α to the carboxyl group. For instance, the amino group may be β, δ, ε, ϕ, or γ, or any other position, to the carboxyl group. In some embodiments Z is γ-aminobutyric acid.

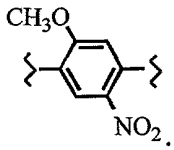
[0074] Certain other specific embodiments of the invention include, without limitation,

Acyl-$CH_2CH_2CH_2$-O-Ph-CH_2-G-NH-C(O)-CH_2I,

Acyl-$CH_2CH_2CH_2$-O-Ph-CH_2-A-NH-C(O)-CH_2I,

Acyl-$CH_2CH_2CH_2$-O-Ph-CH_2-γ-aminobutyric acid-NH-C(O)-CH_2I, and

Acyl-$CH_2CH_2CH_2$-O-Ph-CH_2-V-NH-C(O)-CH_2I,

where Ph is  .

II. Determination of Levels of Expression

[0075] In another aspect, the invention provides for a method for simultaneously identifying and determining the levels of expression of cysteine-containing proteins in normal and perturbed cells, comprising:

a) preparing a first protein sample or a first peptide sample from the normal cells;

b) reacting the first protein sample or the first peptide sample with a reagent of Formula II or III:

(II) Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where:

A is an integer from 0 to 12;

X is selected from the group consisting of an amide bond of formula -C(O)-NR-, a carbonyl of formula -C(O)-, and an amino acid sequence comprising between 0 to 50 amino acids, where R is hydrogen or lower alkyl;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is selected from the group consisting of an amide bond of formula -$(CH_2)_B$-C(O)-NR-, an amide bond of formula -$(CH_2)_B$-NR-C(O)-, and an amino acid sequence comprising between 0 to 10 amino acids,

where R is hydrogen or lower alkyl, and

where B is an integer from 0 to 20;

alk is straight or branched chain of alkylene comprising between 0 and 20 carbon atoms;

Ph is a phenyl group optionally substituted with one or more electron withdrawing groups ortho or para to the -CH_2- group;

Link is selected from the group consisting of -$(CH_2)_C$-I, -$(CH_2)_D$-CH(-$(CH_2)_E CH_3$)-$(CH_2)_F$-X-I, Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide

where C, D, E, and F are each independently an integer from 0 to 20;

Epitope Tag Site is a sequence of amino acids,

where when A is two or more, the amino acid sequence of each Epitope Tag Site can be the same or different; and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme;

c) preparing a second protein sample or a second peptide sample from the perturbed cells;

d) reacting the second protein sample or the second peptide sample of step c) with a second reagent of Formula II or III:

(II) Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where:

A is an integer from 0 to 12;

X is selected from the group consisting of an amide bond of formula -C(O)-NR-, a carbonyl of formula -C(O)-, and an amino acid sequence comprising between 0 to 50 amino acids, where R is hydrogen or lower alkyl;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is selected from the group consisting of an amide bond of formula -$(CH_2)_B$-C(O)-NR-, an amide bond of formula -$(CH_2)_B$-NR-C(O)-, and an amino acid sequence comprising between 0 to 10 amino acids,

where R is hydrogen or lower alkyl, and

where B is an integer from 0 to 20;

alk is straight or branched chain of alkylene comprising between 0 and 20 carbon atoms;

Ph is a phenyl group optionally substituted with one or more electron withdrawing groups ortho or para to the -CH_2- group;

Link is selected from the group consisting of -$(CH_2)_C$-I, -$(CH_2)_D$-CH(-$(CH_2)_E CH_3$)-$(CH_2)_F$-X-I, Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide

where C, D, E, and F are each independently an integer from 0 to 20;

Epitope Tag Site is a sequence of amino acids,

where when A is two or more, the amino acid sequence of each Epitope Tag Site can be the same or different; and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme,

such that the molecular weight of the first reagent and the molecular weight of the second reagent are different by an integer multiple of 14 atomic mass units;

e) combining the reacted the first and the second protein samples or the reacted the first and the second peptide sample from steps b) and d);

f) subjecting the combined protein samples or the combined peptide samples from step e) to proteolysis at a site on the protein samples or at a site on the peptide samples, the site being other than the Protease Cleavage Site;

g) subjecting the proteolyzed combined protein samples or the proteolyzed peptide samples from step f) to an affinity chromatography system comprising a second amino acid sequence attached to a solid, thereby forming bound proteins and non-bound proteins,

where the Epitope Tag Site of the reagent and the second amino acid sequence bind with high specificity to each other;

h) eluting the non-bound proteins from the affinity chromatography system;

i) subjecting the affinity chromatography system from step h) to a protease specific for the Protease Cleavage Site, thereby forming a cleaved protein mixture;

j) eluting the cleaved protein mixture from the affinity chromatography system of step i);

k) isolating the eluted protein mixture obtained from step j);

l) subjecting the eluted protein mixture from step k) to chromatographic separation, followed by mass analysis;

m) comparing the results of step l) to:

1) determining the ratio of amounts of compounds in the two samples, where the molecular weights thereof are separated by an integer multiple of 14 atomic mass units; and

2) comparing the results obtained for each compound to protein databases containing chromatographic and molecular weight correlations.

[0076] In another aspect, the invention provides for a method for simultaneously identifying and determining the levels of expression of cysteine-containing proteins in normal and perturbed cells, comprising:

a) preparing a first protein sample or a first peptide sample from the normal cells;

b) subjecting the first protein sample or the first peptide sample from step a) to proteolysis;

c) reacting the proteolyzed first protein sample or the proteolyzed first peptide sample with a reagent of Formula II or III:

(II) Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where:

A is an integer from 0 to 12;

X is selected from the group consisting of an amide bond of formula -C(O)-NR-, a carbonyl of formula -C(O)-, and an amino acid sequence comprising between 0 to 50 amino acids, where R is hydrogen or lower alkyl;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is selected from the group consisting of an amide bond of formula -$(CH_2)_B$-C(O)-NR-, an amide bond of formula -$(CH_2)_B$-NR-C(O)-, and an amino acid sequence comprising between 0 to 10 amino acids,

where R is hydrogen or lower alkyl, and

where B is an integer from 0 to 20;

alk is straight or branched chain of alkylene comprising between 0 and 20 carbon atoms;

Ph is a phenyl group optionally substituted with one or more electron withdrawing groups ortho or para to the $-CH_2-$ group;

Link is selected from the group consisting of $-(CH_2)_C$-I, $-(CH_2)_D-CH(-(CH_2)_ECH_3)-(CH_2)_F$-X-I, Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide

where C, D, E, and F are each independently an integer from 0 to 20;

Epitope Tag Site is a sequence of amino acids,

where when A is two or more, the amino acid sequence of each Epitope Tag Site can be the same or different; and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme;

d) preparing a second protein sample or a second peptide sample from the perturbed cells;

e) subjecting the second protein sample or the second peptide sample from step d) to proteolysis;

f) reacting the proteolyzed second protein sample or the proteolyzed second peptide sample of step e) with a second reagent of Formula II or III:

(II) Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where:

A is an integer from 0 to 12;

X is selected from the group consisting of an amide bond of formula -C(O)-NR-, a carbonyl of formula -C(O)-, and an amino acid sequence comprising between 0 to 50 amino acids, where R is hydrogen or lower alkyl;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is selected from the group consisting of an amide bond of formula -$(CH_2)_B$-C(O)-NR-, an amide bond of formula -$(CH_2)_B$-NR-C(O)-, and an amino acid sequence comprising between 0 to 10 amino acids,

where R is hydrogen or lower alkyl, and

where B is an integer from 0 to 20;

alk is straight or branched chain of alkylene comprising between 0 and 20 carbon atoms;

Ph is a phenyl group optionally substituted with one or more electron withdrawing groups ortho or para to the -CH2- group;

Link is selected from the group consisting of -$(CH_2)_C$-I, -$(CH_2)_D$-CH(-$(CH_2)_E CH_3$)-$(CH_2)_F$-X-I, Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide

where C, D, E, and F are each independently an integer from 0 to 20;

Epitope Tag Site is a sequence of amino acids,

where when A is two or more, the amino acid sequence of each Epitope Tag Site can be the same or different; and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme,

such that the molecular weight of the first reagent and the molecular weight of the second reagent are different by an integer multiple of 14 atomic mass units;

g) combining the reacted the first and the second protein samples or the reacted the first and the second peptide sample from steps c) and f);

h) subjecting the combined protein samples or the combined peptide samples from step e) to proteolysis at a site on the protein samples or at a site on the peptide samples, the site being other than the Protease Cleavage Site;

i) subjecting the proteolyzed combined protein samples or the proteolyzed peptide samples from step f) to an affinity chromatography system comprising a second amino acid sequence attached to a solid, thereby forming bound proteins and non-bound proteins,

where the Epitope Tag Site of the reagent and the second amino acid sequence bind with high specificity to each other;

j) eluting the non-bound proteins from the affinity chromatography system;

k) subjecting the affinity chromatography system from step j) to a protease specific for the Protease Cleavage Site, thereby forming a cleaved protein mixture;

l) eluting the cleaved protein mixture from the affinity chromatography system of step k);

m) isolating the eluted protein mixture obtained from step l);

n) subjecting the eluted protein mixture from step m) to chromatographic separation, followed by mass analysis;

o) comparing the results of step n) to:

1) determining the ratio of amounts of compounds in the two samples, where the molecular weights thereof are separated by an integer multiple of 14 atomic mass units; and

2) comparing the results obtained for each compound to protein databases containing chromatographic and molecular weight correlations.

[0077] In certain embodiments, if in step c) in the above method Link is Lys-ε-iodoacetamide, then in step f) Link is Orn-δ-iodoacetamide. Alternatively, if in step c) Link is Orn-δ-iodoacetamide, then in step f) Link is Lys-ε-iodoacetamide. In another embodiment, the Z substituent in the first reagent, *i.e.*, in step c) has a molecular weight that is an integer multiple of 14 atomic mass units different than the Z substituent in the second reagent, *i.e.*, in step f). For example, and without limitation, the Z in the first reagent contains valine whereas the Z in the second reagent contains leucine instead of valine, all the other amino acids in Z, if any, remaining the same between the two reagents.

[0078] In an embodiment, the reagent of step c) is selected from the group consisting of

Acyl-NH-AYPYDVPDYASENLYFQGK-iodoacetamide (SEQ ID NO: 17),

Acyl-NH-AYPYDVPDYASENLYFQGGK-iodoacetamide (SEQ ID NO: 18),

Acyl-NH-AYPYDVPDYASENLYFQGAK-iodoacetamide (SEQ ID NO: 19),

Acyl-NH-AYPYDVPDYASENLYFQG(GABA)K-iodoacetamide (SEQ ID NO: 20),

Acyl-NH-AYPYDVPDYASENLYFQGVK-iodoacetamide (SEQ ID NO: 21),
Acyl-NH-AYPYDVPDYASENLYFQGR-iodoacetamide (SEQ ID NO: 22),
Acyl-NH-AYPYDVPDYASENLYFQGGR-iodoacetamide (SEQ ID NO: 23),
Acyl-NH-AYPYDVPDYASENLYFQGAR-iodoacetamide (SEQ ID NO: 24),
Acyl-NH-AYPYDVPDYASENLYFQG(GABA)R-iodoacetamide (SEQ ID NO: 25),
Acyl-NH-AYPYDVPDYASENLYFQGVR-iodoacetamide (SEQ ID NO: 26),
Acyl-NH-AYPYDVPDYASENLYFQGOrn-iodoacetamide (SEQ ID NO: 27),
Acyl-NH-AYPYDVPDYASENLYFQGGOrn-iodoacetamide (SEQ ID NO: 28),
Acyl-NH-AYPYDVPDYASENLYFQGAOrn-iodoacetamide (SEQ ID NO: 29),
Acyl-NH-AYPYDVPDYASENLYFQG(GABA)Orn-iodoacetamide (SEQ ID NO: 30), and
Acyl-NH-AYPYDVPDYASENLYFQGVOrn-iodoacetamide (SEQ ID NO: 31).

[0079] Therefore, by way of example only, if the reagent of step c) is
Acyl-NH-AYPYDVPDYASENLYPQGK-iodoacetamide (SEQ ID NO: 32)
the reagent of step f) would be
Acyl-NH-AYPYDVPDYASENLYPQGOrn-iodoacetamide (SEQ ID NO: 33);
and if the reagent of step c) is
Acyl-NH-AYPYDVPDYASENLYPQGOrn-iodoacetamide (SEQ ID NO: 34),
the reagent of step f) would be
Acyl-NH-AYPYDVPDYASENLYPQGK-iodoacetamide (SEQ ID NO: 35).

[0080] Preferably, the reagent of step c) or of step f) reacts with the reactive side chain of one or more of the amino acid residues of the proteins in the first or second protein sample. By "reactive side chain" it is meant the amino acid side chain that is functionalized, or an amino acid side chain that is other than straight chain or branched alkyl. Therefore, the reagent reacts with the first or second protein at an amino acid residue selected from the group consisting of tyrosine, tryptophan, cysteine, methionine, proline, serine, threonine, lysine, histidine, arginine, aspartic acid, glutamic acid, asparagine, and glutamine. In certain embodiments, the reagent reacts at an amino acid residue selected from the group consisting of tyrosine, cysteine, proline, and histidine. In another embodiment, the site of reaction is a cysteine.

[0081] In some embodiments of the present invention, the chromatographic separation of step l) is a multi-dimensional liquid chromatographic separation, which may be a two-dimensional liquid chromatographic separation or a three-dimensional liquid chromatographic separation. The dimensions of the multi-dimensional liquid chromatographic separation are selected from the group consisting of size differentiation, charge differentiation, hydrophobicity, hydrophilicity, and polarity. In some embodiments, at least one dimension of the multi-dimensional liquid chromatographic separation is separation using size differentiation. Embodiments of the invention include those in which one dimension of the multi-dimensional liquid chromatographic separation is separation using charge differentiation. In other embodiments, one dimension of the multi-dimensional liquid chromatographic separation is separation using hydrophobicity or hydrophilicity.

[0082] In another embodiment the mass analysis of step n) is a multi-dimensional mass analysis, which may be a two-dimensional mass analysis (i.e., tandem mass spectrometry).

[0083] It is well-known in the art to separate fragments of a solution using chromatography and, in tandem thereto, analyze the mass spectra of each fragment. The technique is formally known in the art as LC-MS or LC-MS/MS analysis. Multi-dimensional chromatography is also well-known in the art, where multiple columns are used in tandem, or the same column is packed with segments of different material that can separate the sample using different criteria. *See*, for example, Link *et al.*, (1999) or Opitek *et al.* (1997), above. Multi-dimensional mass analysis is a technique known to those skilled in the art as well. In this technique, following an initial ionization, an ion of interest is selected. The selected ion is fragmented and each fragment (known as "daughter ion" or "progeny ion") is now capable of being either analyzed or be subjected to further fragmentation. The technique is fully described in Siuzdak, Mass Spectrometry for Biotechnology, Academic Press, San Diego, CA, 1996, which is incorporated by reference herein in its entirety.

[0084] In certain embodiments, the preparation of proteins from step a) is subjected to orthogonal chromatography before proceeding with the labeling in step c). Orthogonal chromatography is a technique well-known in the art.

[0085] Quantitative relative amounts of proteins in one or more different samples containing protein mixtures (*e.g.*, biological fluids, cell or tissue lysates, *etc.*) can be determined using chemically similar, affinity tagged and differentially labeled reagents to affinity tag and differentially label proteins in the different samples. The label may be differentiated by having additional methylene groups, which would result in the mass of the two labels be different by an integer multiple of 14.

[0086] In this method, each sample to be compared is treated with a different labeled reagent to tag certain proteins therein with the affinity label. The treated samples are then combined, preferably in equal amounts, and the proteins in the combined sample are enzymatically digested, if necessary, to generate peptides. Some of the peptides are affinity tagged and in addition tagged peptides originating from different samples are differentially labeled. As described above, affinity labeled peptides are isolated, released from the capture reagent and analyzed by (LC/MS). Peptides characteristic of their protein origin are sequenced using $(MS)^n$ techniques allowing identification of proteins in the samples. The relative amounts of a given protein in each sample is determined by comparing relative abundance of the ions generated from any differentially labeled peptides originating from that protein. The method can be used to assess relative amounts of known proteins in different samples. The method is described in U.S. Patent No. 5,538,897, issued July 23, 1996, to Yates *et al.*, which is incorporated herein by reference in its entirety, including any drawings.

[0087] Further, since the method does not require any prior knowledge of the type of proteins that may be present in the samples, it can be used to identify proteins which are present at different levels in the samples examined. More specifically, the method can be applied to screen for and identify proteins which exhibit differential expression in cells, tissue or biological fluids. It is also possible to determine the absolute amount of specific proteins in a complex mixture. In this case, a known amount of internal standard, one for each specific protein in the mixture to be quantified, is added to the sample to be analyzed. The internal standard is an affinity tagged peptide that is identical in chemical structure to the affinity tagged peptide to be quantified except that the internal standard is differentially labeled, either in the peptide or in the affinity tagged portion, to distinguish it from the affinity tagged peptide to be quantified. The internal standard can be provided in the sample

to be analyzed in other ways. For example, a specific protein or set of proteins can be chemically tagged with a labeled affinity tagging reagent. A known amount of this material can be added to the sample to be analyzed. Alternatively, a specific protein or set of proteins may be labeled with additional methylene groups and then derivatized with an affinity tagging reagent.

[0088] Also, it is possible to quantify the levels of specific proteins in multiple samples in a single analysis (multiplexing). For example, a set of five different samples can be reacted with one of SEQ ID NO:27 - SEQ ID NO:31, then follow with subsequent steps as described herein. In this case, affinity tagging reagents used to derivatize proteins present in different affinity tagged peptides from different samples can be selectively quantified by mass spectrometry. This may be achieved by using reagents whose molecular mass varies from one sample to another by an integer multiple of 14. So, for example, the Link group in one reagent may feature ornithine whereas the Link group in another reagent may feature arginine or lysine. Similarly, the Z groups in the different reagent may vary such that the molecular mass of the reagent varies by an integer multiple of 14. It is also understood that other amino acids may also be featured. For example, the lighter reagent may have valine whereas the heavier reagent may feature leucine or isoluecine in its stead. The same would be true for having asparagine in the lighter reagent and glutamine in the heavier reagent, or aspartic acid in the lighter reagent and glutamic acid in the heavier reagent.

[0089] In this aspect of the invention, the method provides for quantitative measurement of specific proteins in biological fluids, cells or tissues and can be applied to determine global protein expression profiles in different cells and tissues. The same general strategy can be broadened to achieve the proteome-wide, qualitative and quantitative analysis of the state of modification of proteins, by employing affinity reagents with differing specificity for reaction with proteins. The method and reagents can be used to identify low abundance proteins in complex mixtures and can be used to selectively analyze specific groups or classes of proteins such as membrane or cell surface proteins, or proteins contained within organelles, sub-cellular fractions, or biochemical fractions such as immunoprecipitates. Further, these methods can be applied to analyze differences in expressed proteins in different cell states. For example, the methods and reagents herein can

be employed in diagnostic assays for the detection of the presence or the absence of one or more proteins indicative of a disease state, such as cancer.

[0090] The methods described herein can also be applied to determine the relative quantities of one or more proteins in two or more protein samples. The proteins in each sample are reacted with affinity tagging reagents which are substantially chemically identical but differentially labeled. The samples are combined and processed as one. The relative quantity of each tagged peptide which reflects the relative quantity of the protein from which the peptide originates is determined by the integration of the respective mass peaks by mass spectrometry.

[0091] The methods described herein can be applied to the analysis or comparison of multiple different samples. Samples that can be analyzed by methods of this invention include cell homogenates; cell fractions; biological fluids including urine, blood, and cerebrospinal fluid; tissue homogenates; tears; feces; saliva; lavage fluids such as lung or peritoneal lavages; mixtures of biological molecules including proteins, lipids, carbohydrates and nucleic acids generated by partial or complete fractionation of cell or tissue homogenates.

[0092] The methods described herein employ MS and $(MS)^n$ methods. While a variety of MS and $(MS)^n$ are available and may be used in these methods, Matrix Assisted Laser Desorption Ionization MS (MALDI/MS) and Electrospray ionization MS (ESI/MS) methods are preferred.

III. Proteomic Analysis

[0093] Another aspect of the present invention relates to a method for proteomic analysis, comprising:

a) preparing a protein sample or a peptide sample from cells;

b) reacting the protein sample or the peptide sample with a reagent of the formula:

Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Link

where:

A is an integer from 1 to 12;

X is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or X is an amino acid sequence comprising between 0 to 50 amino acids;

Y is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Y is an amino acid sequence comprising between 0 to 50 amino acids;

Z is an amide bond of formula -C(O)-NR-, where R is hydrogen or lower alkyl, or Z is an amino acid sequence comprising between 0 to 10 amino acids;

Link is selected from the group consisting of Lys-ε-iodoacetamide, Arg-δ-iodoacetamide, and Orn-δ-iodoacetamide;

Epitope Tag Site is a sequence of amino acids, and

Protease Cleavage Site is a sequence of amino acids that is a cleavage site for a highly specific protease enzyme;

c) subjecting the reacted proteins or peptides from step b) to proteolysis at a site on the protein samples or at a site on the peptide samples, the site being other than the Protease Cleavage Site;

d) subjecting the proteolyzed reacted proteins or the proteolyzed reacted peptides from step c) to an affinity chromatography system comprising a second amino acid sequence attached to a solid support, thereby forming bound proteins and non-bound proteins,

where the Epitope Tag Site of the reagent and the second amino acid sequence bind with high specificity to each other;

e) eluting the non-bound proteins from the affinity chromatography system;

f) subjecting the affinity chromatography system from step e) to a protease specific for the Protease Cleavage Site, thereby forming a cleaved protein mixture;

g) eluting the cleaved protein mixture from the affinity chromatography system of step f);

h) isolating the cleaved protein mixture obtained from step g);

i) subjecting the cleaved protein mixture from step h) to chromatographic separation, followed by mass analysis;

j) comparing the results of step i) to:

1) determine the ratio of amounts of compounds in the sample separated by a molecular weight of 14 atomic mass units; and

2) identify the various modified proteins by comparing the results obtained for each modified protein to protein databases containing chromatographic and molecular weight correlations.

[0094] The term "proteomic analysis" refers to identifying the proteome of a cell. The "proteome" of a cell is the collection of all the proteins expressed by the cell at the time the proteomic analysis is undertaken. It is understood that, unlike the genome of a cell, which is invariable, the proteome of a cell varies depending on many factors, including the age of the cell, the environmental conditions surrounding the cell, and the position of the cell in its life cycle.

[0095] In the above methods, the reagent reacts with the reactive side chain of one or more of the amino acid residues of the first or second protein. Therefore, the reagent reacts with the protein at an amino acid residue selected from the group consisting of tyrosine, tryptophan, cysteine, methionine, proline, serine, threonine, lysine, histidine, arginine, aspartic acid, glutamic acid, asparagine, and glutamine. In certain embodiments, the reagent reacts at an amino acid residue selected from the group consisting of tyrosine, cysteine, proline, and histidine. In another preferred embodiment, the site of reaction is a cysteine.

[0096] In some embodiments of the present invention, the chromatographic separation of step i) is a multi-dimensional liquid chromatographic separation, which may be a two-dimensional liquid chromatographic separation or a three-dimensional liquid chromatographic separation. The dimensions of the multi-dimensional liquid chromatographic separation are selected from the group consisting of size differentiation, charge differentiation, hydrophobicity, hydrophilicity, and polarity. In some embodiments, at least one dimension of the multi-dimensional liquid chromatographic separation is separation using size differentiation. Embodiments of the invention include those in which one dimension of the multi-dimensional liquid chromatographic separation is separation using charge differentiation. In other embodiments, one dimension of the multi-dimensional liquid chromatographic separation is separation using hydrophobicity or hydrophilicity.

[0097] In another embodiment the mass analysis of step i) is a multi-dimensional mass analysis, which more preferably, may be a two-dimensional mass analysis.

[0098] In certain embodiments, the preparation of proteins from step a) is subjected to orthogonal chromatography before proceeding with the labeling in step b).

[0099] In one aspect, the invention provides a mass spectrometric method for identification and quantification of one or more proteins in a complex mixture which employs affinity labeled reagents in which the Link group is a group that selectively reacts with certain groups that are typically found in peptides (*e.g.*, sulfhydryl, amino, carboxy, homoserine, or lactone groups). One or more affinity labeled reagents with different Link groups are introduced into a mixture containing proteins and the reagents react with certain proteins to tag them with the affinity label. It may be necessary to pretreat the protein mixture to reduce disulfide bonds or otherwise facilitate affinity labeling. After reaction with the affinity labeled reagents, proteins in the complex mixture are cleaved, *e.g.*, enzymatically, into a number of peptides. This digestion step may not be necessary, if the proteins are relatively small. Peptides that remain tagged with the affinity label are isolated by an affinity isolation method, *e.g.*, affinity chromatography, via their selective binding to the capture reagent. Isolated peptides are released from the capture reagent by displacement of the Epitope Tag Site or cleavage of the linker, and released materials are analyzed by liquid chromatography/mass spectrometry (LC/MS). The sequence of one or more tagged peptides is then determined by $(MS)^n$ techniques. At least one peptide sequence derived from a protein will be characteristic of that protein and be indicative of its presence in the mixture. Thus, the sequences of the peptides typically provide sufficient information to identify one or more proteins present in a mixture.

<u>Quantitative Proteome Analysis</u>

[0100] The method comprises the following steps:

[0101] <u>Reduction</u>. Disulfide bonds of proteins in the sample and reference mixtures are chemically reduced to free SH groups. The preferred reducing agent is tri-*n*-butylphosphine which is used under standard conditions. Alternative reducing agents include mercaptoethanol, 2-methylthioethanol, 2-methylthio-1-hexanol, and dithiothreitol. If required, this reaction can be performed in the presence of solubilizing agents including high concentrations of urea and detergents to maintain protein solubility. The reference and

sample protein mixtures to be compared are processed separately, applying identical reaction conditions.

[0102] Derivatization of SH groups with an affinity tag. Free SH groups of the sample protein are derivatized with a reagent of the invention. The reagent reacts with the free SH group through the Link group.

[0103] Each sample is derivatized with a different reagent having a different mass. Derivatization of SH groups is preferably performed under slightly basic conditions (pH 8.5) for 90 min at about room temperature. For the quantitative, comparative analysis of two samples, one sample each (termed "reference sample" and "sample") are derivatized with two different reagents, whose molecular mass differs by an integer multiple of 14. For the comparative analysis of several samples one sample is designated a reference to which the other samples are related.

[0104] Combination of labeled samples. After completion of the affinity tagging reaction defined aliquots of the samples labeled with different reagents are combined and all the subsequent steps are performed on the pooled samples. Combination of the differentially labeled samples at this early stage of the procedure eliminates variability due to subsequent reactions and manipulations. Preferably equal amounts of each sample are combined.

[0105] Removal of excess affinity tagged reagent. Excess reagent is adsorbed, for example, by adding an excess of SH-containing beads to the reaction mixture after protein SH groups are completely derivatized. Beads are added to the solution to achieve about a 5-fold molar excess of SH groups over the reagent added and incubated for 30 min at about room temperature. After the reaction the beads are removed by centrifugation.

[0106] Protein digestion. The proteins in the sample mixture are digested, typically with trypsin. Alternative proteases are also compatible with the procedure as in fact are chemical fragmentation procedures. In cases in which the preceding steps were performed in the presence of high concentrations of denaturing solubilizing agents, the sample mixture is diluted until the denaturant concentration is compatible with the activity of the proteases used. This step may be omitted in the analysis of small proteins.

[0107] Affinity isolation of the affinity tagged peptides by interaction with a capture reagent. The tagged peptides are isolated on anti-HA antibodies-agarose. After

digestion the pH of the peptide samples is lowered to 6.5 and the tagged peptides are immobilized on beads coated with anti-HA. The beads are extensively washed. The last washing solvent includes 10% methanol to remove residual SDS.

[0108] Release of the captured peptides with specific protease. A solution of TEV in TRIS at pH 7.5 is added to the column and digestion is allowed to proceed. The bound peptides are cleaved from the column by incubation at 30 °C for 6 hours.

[0109] Analysis of the isolated, derivatized peptides by μLC-$(MS)^n$ or CE-$(MS)^n$ with data dependent fragmentation. Methods and instrument control protocols well-known in the art and described, for example, in Ducret *et al.* (1998); Figeys and Aebersold (1998); Figeys *et al.* (1996); or Haynes *et al.* (*Electrophoresis* 19:939-945 (1998)) are used.

[0110] In this last step, both the quantity and sequence identity of the proteins from which the tagged peptides originated can be determined by automated multistage MS. This is achieved by the operation of the mass spectrometer in a dual mode in which it alternates in successive scans between measuring the relative quantities of peptides eluting from the capillary column and recording the sequence information of selected peptides. Peptides are quantified by measuring in the MS mode the relative signal intensities for pairs of peptide ions of identical sequence that are tagged with the lighter or heavier forms of the reagent, respectively, and which therefore differ in mass by the mass differential encoded within the affinity tagged reagent. Peptide sequence information is automatically generated by selecting peptide ions of a particular mass-to-charge (*m/z*) ratio for collision-induced dissociation (CID) in the mass spectrometer operating in the $(MS)^n$ mode. (Link *et al.* *Electrophoresis* 18:1314-1334 (1997); Gygi *et al. Nature Biotechnol* 17:994-999 (1999); Gygi *et al.*, *Cell Biol* 19:1720-1730 (1999)). The resulting CID spectra are then automatically correlated with sequence databases to identify the protein from which the sequenced peptide originated. Combination of the results generated by MS and $(MS)^n$ analyses of affinity tagged and differentially labeled peptide samples therefore determines the relative quantities as well as the sequence identities of the components of protein mixtures in a single, automated operation.

[0111] This method can also be practiced using other affinity tags and other protein reactive groups, including amino reactive groups, carboxyl reactive groups, or groups that react with homoserine lactones.

[0112] The approach employed herein for quantitative proteome analysis is based on two principles. First, a short sequence of contiguous amino acids from a protein contains sufficient information to uniquely identify that protein. Protein identification by $(MS)^n$ is accomplished by correlating the sequence information contained in the CID mass spectrum with sequence databases, using sophisticated computer searching algorithms (Yates, III *et al.* U.S. Patent 5,538,897). Second, pairs of peptides tagged with lighter and heavier Link groups or Z groups, respectively, are chemically similar and therefore serve as mutual internal standards for accurate quantification. The MS measurement readily differentiates between peptides originating from different samples, representing for example different cell states, because of the difference between the distinct reagents attached to the peptides. The ratios between the intensities of the differing weight components of these pairs or sets of peaks provide an accurate measure of the relative abundance of the peptides (and hence the proteins) in the original cell pools.

[0113] Specifically, the peptide labeling moiety consists of a lysine residue modified with an iodoacetamido functional group on the ε-amino side chain. The synthetic chemistry necessary for this modification reaction is readily available in the literature. The synthetic peptides contain two additional motifs: a peptide epitope tag for high affinity purification; and a highly specific protease site for releasing the affinity purified labeled peptides from the affinity matrix. In addition, these synthetic peptides can readily be prepared as isoforms of two different masses by the simple expedient of using an ornithine in place of lysine to introduce a 14 mass unit difference in the carboxyl terminal acid.

[0114] Examples of the reagents (SEQ ID NO: 36 and SEQ ID NO: 37) are thus:

Ala-[Tyr-Pro-Tyr-Asp-Val-Pro-Asp-Tyr-Ala]-Ser-(Glu-Asn-Leu-Tyr-Phe-Gln-Gly)-Lys---Iodoacetamide

(Epitope Tag Site) (Protease Cleavage Site)

Ala-[Tyr-Pro-Tyr-Asp-Val-Pro-Asp-Tyr-Ala]-Ser-(Glu-Asn-Leu-Tyr-Phe-Gln-Gly)-Orn---Iodoacetamide

[0115] The peptide sequence in the square brackets is an Epitope Tag Site and the sequence in parentheses is a Protease Cleavage Site. In the case shown here, the peptide sequence YPYDVPDYA (SEQ ID NO: 38) is an influenza hemagglutinin (HA) epitope tag. This part of the reagent could be replaced by any other epitope tag, or multiple copies of a single tag for higher efficiency purification, or parallel copies of different tags for higher specificity purification. Examples of other Epitope Tag Sites include Flag, His-6, and c-myc.

[0116] The protease cleavage site shown here is that of TEV protease, which is commercially available. This enzyme has been shown to cleave at only one protein site in the entire yeast genome, thus indicating that the enzyme is highly specific for an extremely rare sequence. This part of the reagent could be replaced by any other highly specific protease cleavage site, either commercially available, such as Factor Xa, or Pharmacia Prescission Enzyme, or one that is newly discovered. The amino acid indicated in bold is used to provide a site of attachment for the iodoacetamide group, hence we have used lysine which contains an ε-amino side chain that is suitable for the purpose. This amino acid is also used to introduce a differential mass between the two reagents, and this can be readily accomplished by using ornithine in place of lysine. Ornithine is commercially available and differs from lysine only by the presence of one additional methyl group, which makes it 14 amu (atomic mass unit) heavier than lysine. Arginine is also commercially available and its molecular weight is 28 amu (*i.e.*, 2 x 14) heavier than lysine. This part of the reagent could be replaced with any other amino acid or similar molecule that provided an attachment site for the iodoacetamide group. Finally, the integral difference of 14 amu could be further enhanced by the choice of two amino acids differing by 14 amu (*e.g.*, valine and leucine) in the Z portion of the peptide labeling moiety.

Qualitative Proteome Analysis

[0117] In addition to the above methods, the methods of the invention may be used to determine the proteomic differences in an organism or cell based on the change in the cell's environmental condition. Thus, for example, one may compare the proteome of the cells of two plants of the same species, one having encountered high salt concentrations and

the other low salt concentrations, thereby determining the effect of salt concentration on the plant's proteome.

[0118] It is also within the scope of the present invention that the two modes of analysis discussed herein, *i.e.*, the qualitative and quantitative proteome analyses, are exercised in conjunction with each other. Thus, by way of example only, one may compare the proteome of the cells of two plants of the same species, one having encountered higher temperatures than the other, thereby not only determining the effect of heat on the proteome in terms of which proteins are expressed, but also determining the effect of heat on the level of expression of each protein of interest.

[0119] In practicing the present invention to achieve the above end, one may use a number of different compounds of the present invention, having different masses (yet all within an integer multiple of 14 from each other), and mark different proteins of the cells with the different reagents. By applying the multidimensional LC/MS techniques described herein, one is able to determine which proteins, and to what extent, are expressed in the cells.

IV. Fusion Proteins

[0120] Another aspect of the invention relates to a process for preparing a fusion protein of Formula IV or V:

(IV) Protein-Acyl-N-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-[Lys-δ-N-iodoacetamide]

(V) Protein-Acyl-NH-X-alk-O-Ph-CH_2-Z-Link

where A, X, Y, Z, alk, Ph, Link, Epitope Tag Site, and Protease Cleavage Site are as defined herein

comprising,

a) preparing a fusion protein sample of Formula II or III from cells

(II) Protein-Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Orn-δ-$NHCOCH_2$

(III) Acyl-NH-X-alk-O-Ph-CH_2-Z-$NHCOCH_2$

b) reacting the protein sample with a Link or with iodoacetamide.

[0121] In another aspect, the invention relates to a process for preparing a fusion protein of Formula VI:

(VI) Protein-Acyl-N-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-[Lys-δ-N-iodoacetamide]

where A, X, Y, Z, alk, Ph, Link, Epitope Tag Site, and Protease Cleavage Site are as defined herein

comprising,

a) preparing a fusion protein sample of Formula VII from cells

(VII) Protein-Acyl-NH-X-[Epitope Tag Site]$_A$-Y-[Protease Cleavage Site]-Z-Lys-δ-$NHCOCH_2$

b) reacting the protein sample with iodoacetamide.

[0122] Markers that are useful in plant breeding, genetics, and diagnostics are disclosed in U.S. Provisional Patent Application No. 60/264,226, entitled "Cereal Simple Sequence Repeat Markers," filed on January 26, 2001 (Attorney Docket No. NADII.026PR), which is hereby incorporated by reference in its entirety.

IV. Databases

[0123] Aspects of the invention not only include the chemical compounds and MS data described above, but also include data files (e.g.: databases) corresponding to these compounds and data. For example, the amino acid sequences of the labeled compounds can be created and manipulated *in silico*. These data files can be stored in a conventional computer system on any type of temporary or permanent storage. Examples of such storage include Read Only Memory, Random Access Memory, Hard Disk, Floppy Disk, CD-ROM and the like.

[0124] In addition to data relating to the modified amino acid sequences, aspects of the invention include data files of the MS data itself. A data file of, for example, a cell that

has been subjected to high salt conditions, can be stored to a database and thereafter compared to other data files of cells having different treatments. Thus, aspects of the invention contemplate analyzing the differences between organisms or cells by comparing MS data gathered from the methods described above.

Examples

[0125] Examples are provided below to illustrate different aspects and embodiments of the present invention. These examples are not intended in any way to limit the disclosed invention. Rather, they illustrate the compounds and the methodology by which the protein analysis of the invention may be practiced.

[0126] The following proteins and reagents were purchased from Sigma, St. Louis, MO, USA: rabit glyceraldehydes-3-phosphate dehydrogenase, E.Coli β-galactosidase, rabbit phosphorylase b, chicken ovalbumin, bovine β-lactoglobulin, bovine a-lactalbumin, bovine serum albumin, dimethylformamide (DMF), Iodoacetic anhydride, Urea, tris-hydrochloride, acid washed glass beads, and diisopropylethylamine (DIEA). Tributyl phosphine was purchased from BioRad (Hercules, CA). Synthetic peptides were custom made by QCB/Biosource International (Hopkinton, MA). HA affinity matrix and Lys-C were from Roche Diagnostics (Indianapolis, IN), and PreScission protease was from Amersham Pharmcia Biotech (Uppsala, Sweden). HPLC grade acetonitrile (ACN) and HPLC grade methanol was purchased from Fischer Scientific (Fair Lawn, NJ). Yeast extract were products of BD Biosciences (Sparks, MD). Heptaflourobutyric acid (HFBA) was obtained from Pierce (Rockford, IL). SPEC Plus PT C18 solid phase extraction pipette tips were purchased from Ansys Diagnostics (Lake Forest, CA). Glacial acetic acid was purchased from Malinckrodt Baker Inc. (Paris, KY).

Example 1: Synthesis of peptide labeling moiety (or peptide encoded tags, "PEPTags")

[0127] A pair of PEPTags, described generally above, was synthesized from peptides with following sequences: Ac-AYPYDVPDYASENLYFQGK (SEQ ID NO: 39) and AYPYDVPDYASENLYFQGOrn (SEQ ID NO: 40). In dry DMF containing excess (2-3 molar equivalents) DIEA, each of the peptides was mixed with two molar equivalents of

iodoacetic anhydride for 10 min at room temperature under N_2 gas, to give Lys-PEPTag and Orn-PEPTag, respectively. The reaction was terminated by adding acetic acid. Solvent was removed by vacuum centrifugation, and the product was purified by reverse-phase FPLC, and analyzed by MALDI MS (TofSpec 2E, Micromass, Beverly, MA) and ESI MS/MS (API 3, PE Sciex, Foster City, CA).

[0128] In order to demonstrate that the mass spectrometric ionization efficiency of the two synthesized peptide tags was essentially equal, the two products were mixed in different ratios and analysed by LC-MS. The ratio of the measured peak areas gave the data shown in the following table.

Amount of tag1 (pmol)	Amount of tag2 (pmol)	Calculated ratio	Measured ratio
30	3	10:1	11.95:1
15	3	5:1	5.19:1
7.5	3	2.5:1	2.70:1
3.75	3	1.25:1	0.97:1
1.875	3	0.625:1	0.64:1
0.375	3	0.125:1	0.11:1

Example 2: PEPTag qualitative protein analysis: simplification of complex mixtures**[0129]** We tested the PEPTag method, described generally herein, on Bovine Serum Albumin (BSA). 200 µL BSA (0.25 mg/mL) was denatured and reduced in a solution containing 0.1% SDS, 5 mM tributyl phosphine and 50 mM Tris buffer (pH 8.5) for 3 min at 100 °C and for 1 hour at 37 °C. The side chains of cysteinyl residues were derivatized with a tenfold molar excess of Lys- PEPTag. Tagged protein was digested by trypsin overnight at 37 °C. Trypsin activity was quenched with trypsin inhibitor and the peptide mixture bound to anti-HA affinity matrix for 2 hours at 4 °C. The anti-HA resin with bound peptides was washed in equilibration -buffer (20mM Tris, pH 7.5; 0.1 M NaCl; 0.1mM EDTA), 3 X 10 min. at 4 °C. The bound peptides were cleaved from the matrix by incubation with TEV protease for 6 hours at 30 °C. The cleaved peptides were analyzed by either Matrix Assisted

Laser Desorption Ionization Mass Spectrometry (MALDI MS), or separated and analyzed by μLC-MS/MS. Using the Sequest database searching algorithm (Yates, III *et al.* U.S. Patent 5,538,897), the resulting MS/MS spectra were correlated with the sequence database.

[0130] The sequence of bovine serum albumin is shown below:

SW:ALBU_BOVIN P02769 bos taurus (bovine). serum albumin precursor. 12/1998
[MASS=69293]
MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA FSQYLQQCPF
DEHVKLVNEL TEFAKTCVAD ESHAGCEK**SL HTLFGDELCK** VASLRETYGD MADCCEK**QEP**
ERNECFLSHK DDSPDLPKLK PDPNTLCDEF KADEKKFWGK YLYEIARRHP YFYAPELLYY
ANKYNGVFQE CCQAEDKGAC LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA
RLSQKFPKAE FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAK**YICDN QDTISSK**LKE
CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL GSFLYEYSRR
HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL KHLVDEPQNL IKQNCDQFEK
LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS RSLGKVGTRC **CTKPESERMP CTEDYLSLIL**
NRLCVLHEKT PVSEKVTKCC TESLVNR**RPC FSALTPDETY VPK**AFDEK**LF TFHADICTLP**
DTEKQIKKQT ALVELLKHKP KATEEQLK**TV MENFVAFVDK** CCAADDKEAC FAVEGPKLVV STQTALA
>average mass = **69294**, pI = **5.82**

[0131] Cysteine-containing peptides indicated in bold-underline are those detected in the experiment described in example 2. The protein is successfully identified from each peptide tandem MS spectra, and the complex total tryptic mixture of peptides is considerably simplified. The peptides are shown in more detail in the table below, with C# indicating a peptag-modified cysteine residue.

Position	Mass (MH+)	Peptide sequence
89-100	1363.57	SLHTLFGDELC#K
286-297	1387.50	YIC#DNQDTISSK
139-151	1520.74	LKPDPNTLC#DEFK
510-523	1571.78	C#FSALTPDETYVPK
469-482	1668.96	MPC#TEDYLSLILNR
508-523	1825.08	RPC# FSALTPDETYVPK
123-138	1846.02	NEC#FLSHKDDSPDLPK
529-544	1852.11	LFTFHADIC#TLPDTEK
118-138	2485.68	QEPERNEC#FLSHKDDSPDLPK
461-482	2599.99	CTKPESERMPC#TEDYLSLILN

		R

Example 3: PEPTag quantitative protein analysis: differential labeling

[0132] We tested the PEPTag quantitative strategy on two mixtures containing the same two proteins at different concentrations. Mixture 1 had 500 pmol BSA (0.1 mg/mL) and 400 pmol β-lactoglobulin (0.1 mg/mL) and was reacted with 9 nmol Lys-PEPTag. Mixture 2 had 250 pmol BSA (0.05 mg/mL) and 800 pmol β- lactoglobulin (0.2mg/mL) and was reacted with 9 nmol Orn-PEPTag. Protein denaturation, reduction, tagging, and digestion were the same as described above. The two samples were combined after tryptic digestions, and bound to anti-HA matrix. TEV digestion and MS analysis were as described in Example 2. Peptides were quantified by measuring, in the MS mode, the relative signal intensities for pairs of peptide ions of identical sequence, tagged with Lys or Orn-PEPTags, respectively. The results are shown in Figures 6, 7, and 8 and the following table.

Protein	Peptide sequence identified	Observed ratio	Mean±S.D.	Expecte ratio
Bovine serum albumin	SLHTLFGDELC#K	2.19	2.05±0.10	2.00
	GLVLIAFSQYLQQC#PFDEHVK	1.96		
	GLVLIAFSQYLQQC#PFDEHVKLVNELTEFAK	1.99		
Beta-lactogobulin	VYVEELKPTPEGDGLEILLQKWENDEC#AQKK	0.40	0.46±0.05	0.50
	LSFNPTQLEEQC#HI	0.51		

Example 4: Proteome analysis

A. Perturbed cell sample versus normal cell sample

[0133] A biological sample of interest is subjected to a treatment expected to cause physiological changes, such as treating tissue culture cells with a drug sample. Protein samples are prepared from the normal and perturbed cells. The normal cell protein sample is labeled at all cysteine residues using the first (lysine-based) reagent shown above, and the perturbed cell protein sample is labeled at all cysteine residues using the heavier (ornithine-based) version of the reagent as shown above. The two labeled samples are then combined and protease digested, typically with trypsin, to produce a very complex peptide mixture.

This complex mixture is then passed over an anti-HA tag affinity tag column that retains only those tryptic fragments containing labeled cysteine residues, allowing all other material to be washed away. The peptides are then released from the column by addition of TEV protease, producing a mixture of peptides labeled with either lysine or ornithine attached via an acetamido group.

[0134] This complex mixture is then analyzed using microscale high-performance liquid chromatography–tandem mass spectrometry. Two distinct classes of information are then obtained during the course of a single experiment. Firstly, the relative amounts of each peptide that were produced from the initial normal and perturbed samples are accurately quantified by measuring the ratio of peak areas for a given peak pair differing by 14 amu. Since the two samples have been mixed together very early in the experimental process, variation in sampling handling between the two samples is essentially eliminated as for each pair there is a mutual internal standard present in the same sample. Secondly, the identity of each peptide is determined by tandem mass spectrometric fragmentation and database searching using established methods.

[0135] The result of this experiment is simultaneous peptide identification and relative quantification. Thus, for any experimental perturbation that can be applied to cells, it would be possible to identify which proteins were up and down regulated, and quantify the amount of any change detected.

B. Whole cell analysis

[0136] Another type of experiment is performed using just one of the reagents described above, where massively parallel protein identification is required such as characterizing the proteome of a whole organism or cell type. Using the technique outlined above for enrichment of labeled cysteine containing peptides, the number of proteins that can be identified from a very complex mixture is dramatically increased. This is due to the fact the number of peptides analyzed from each protein, even those of high abundance, is reduced, thus allowing greater coverage of the range of proteins present. This coverage is increased still further by using two-dimensional liquid chromatography prior to tandem mass spectrometry in order to maximize the number of peptides analyzed. It is also possible to

perform a further orthogonal chromatography step prior to labeling, thus increasing the number of peptides identified even more. Using such a system, it is possible to describe the entire proteome of a simple organism in a single experiment.

[0137] The applications of this method are almost limitless. Any biological sample containing proteins benefits from either a complete description of all the proteins present, or a complete description and quantification of changes that occur in response to a physiological stimulus, or both.

[0138] The complete cataloging type of experiment, set forth in Subsection B, above, is best limited to organisms with complete sequences available, although it should be noted that the list now includes humans.

Example 5: Synthesis of affinity peptide encoded tags (APEPTags)

[0139] A pair of APEPTags was synthesized from peptides with following sequences: Ac-AYPYDVPDYASLEVLFQGPK-NH_2 and Ac-AYPYDVPDYASLEVLFQGPOrn-NH_2. In dry DMF containing excess (2-3 molar equivalents) DIEA, each of the peptides was mixed with two molar equivalents of iodoacetic anhydride for 10 min at room temperature under N_2 gas. The reaction was terminated by adding acetic acid. Solvent was removed by vacuum centrifugation, and the product was purified on a Sephasil_Peptide_C18_5μ_ST_4.6/100 column connected to AKTA purifier Amersham Pharmcia Biotech FPLC system (Uppsala, Sweden). Solvent A was 0.01% v/v TFA/H_2O, and solvent B was 0.01 % v/v TFA/ H_2O/90% acetonitrile. A flow rate of 0.8 ml/min was used, with the UV monitored at 280 nm. The gradient was from 0 to 50% B over 35 column volume. The fraction-collected peak was analyzed by MALDI MS (TofSpec 2E, Micromass) with α-cyano-4-hydroxy-cinnamic acid as matrix and by ESI MS/MS (API 3, PE Sciex).

Example 6: Synthesis of immobilized peptide encoded tags (IPEPTags)

[0140] A pair of IPEPTags was synthesized from peptides with following sequences: Sepharose gel-CASASLEVLFQGPK-NH_2 and Sepharose gel-CASASLEVLFQGPOrn-NH_2. Pack two 10 ml empty columns with 2 ml of each gel-

coupled peptide. Drain the storage buffer completely. Rinse the gel bed three times with 5 ml DMF. Add 2 ml DMF with 2 μmol iodoacetic anhydride and 1 μl DIEA into each column. Mix and react at room temperature for 15 min. Drain reagents completely and rinse the gel with 10 X volume of buffer 50 mM tris (pH 8.5) and then store in the same buffer.

Example 7: Growth and Lysis of S. cerevisiae

[0141] Strain BJ5460 was grown to mid log phase (O.D. 0.6) in YPD, centrifuged and washed 1X with buffer (1 M sorbitol, 10 mM KH2PO4, pH 7.5, 50 mM NaCl, 1 mM EDTA). Resuspended cells in buffer, added zymolase (3 mg per 100 OD), and incubate at 30 °C for 45 min. Cells were harvested by centrifugation, wash once and then solubilized in 8 M Urea, 50 mM Tris-HCl pH 8.5 and disrupted in the presence of glass beads on a mixer. The protein concentration was determined by the Bradford assay.

Example 8. APEPTag analysis of protein mixtures

[0142] Protein mixtures were denatured and reduced in a buffer containing 8 M Urea, 10 mM tributyl phosphine and 50 mM Tris buffer (pH 8.5) for 30 min at 50°C. The side chains of cysteinyl residues were derivatized with about 5 fold molar excess of APEPTag. Tagged proteins were dialysis against 50 mM Tris buffer (pH 8.5) for 5 hours and then digested by trypsin overnight at 37 °C. Trypsin activity was quenched with trypsin inhibitor and the peptide mixture bound to anti-HA affinity matrix for 2 hours at 4 °C. The anti-HA resin with bound peptides was washed with 10 volume of equilibration buffer (20mM Tris, pH 7.5; 0.1m NaCl; 0.1mM EDTA), 3 X 10 min. at 4 °C. The bound peptides were cleaved from the matrix by incubation with PreScission protease overnight at 4 °C.

[0143] For APEPTag quantitative strategy, two protein mixtures were denatured, reduced and then labeled differentially with either Lys-APEPTag or Orn-APEPTag. The two mixtures were combined after their dialysis. Protein denaturation, reduction, tagging, dialysis, digestion, affinity binding and were the same as described above.

Example 9. IPEPTag analysis of protein mixtures

[0144] Protein mixtures were denatured and reduced in a buffer containing 8 M Urea, 10 mM tributyl phosphine and 50 mM Tris buffer (pH 8.5) for 30 min at 50 °C. The side chains of cysteinyl residues were derivatized with about 10 fold molar excess of IPEPTag beads. Tagged proteins were digested first by Lys-C in 8M urea for 6 hours and then by trypsin in 2 M urea overnight at 37 °C. The beads with bound peptides were washed with 10 volume of equilibration buffer (20mM Tris, pH 7.5; 0.1m NaCl; 0.1mM EDTA), 3 X 10 min. at 4 °C. The bound peptides were cleaved from the matrix by incubation with PreScission protease overnight at 4 °C.

[0145] For IPEPTag quantitative strategy, two protein mixtures were denatured, reduced and then labeled differentially with either Lys-IPEPTag or Orn-IPEPTag beads. Protein denaturation, reduction, tagging, and digestion were the same as described above. Two batches of beads with bound peptides were combined after digestion, followed by wash and preScission cleavage as described above.

Example 10. Chromatography and Mass Spectrometry

[0146] Each sample was subjected to MudPIT analysis with modifications to the method described by Link et al. A quaternary HP 1100 HPLC pump (Hewlett-Packard, Palo Alto, CA) was interfaced with a Finnigan LCQ ion trap mass spectrometer (Finnigan MAT, San Jose, CA). The tip at the end of the 100 x 365 μm fused silica capillary (J & W Scientific, Folsom, CA) was pulled with a P-2000 laser (Sutter Instruments Co., Novato, CA). The fritless capillary was first packed with 10 cm of 5 μm Zorbax Eclipse XDB-C18 (Hewlett Packard, Palo Alto, CA) and then with 4 cm of 5 μm Partisphere SCX (Whatman, Clifton, New Jersey). The column, was connected to a PEEK micro-cross as described elsewhere, in order to split the flow of the HPLC pump to an effective flow rate of 0.15 -0.25 μL/min and supply a spray voltage of 1.8 W. The Zorbax 4.6 x 30 mm Eclipse XDB C18 column for the off-line fractionation was manufactured by Hewlett Packard, Palo Alto, CA.

[0147] Each sample mixture was loaded onto separate microcolumn for the analysis. After loading the microcapillary column, the column was placed in-line with the system. A fully automated 7-step chromatography run was carried out on each sample. The four buffer solutions used for the chromatography were 5% ACN/0.5% acetic acid/0.02%

HFBA (buffer A), 80% ACN//0.5% acetic acid/0.02% HFBA (buffer B), 250 mM ammonium acetate/5% ACN/0.5% acetic acid/0.02% HFBA (buffer C), and 1.5 M ammonium acetate/5% ACN/0.5% acetic acid/ 0.02% HFBA (buffer D). The first step of 80 min consisted of a 70 min gradient from 0 to 80% buffer B and a 10 min hold at 80% buffer B. The next 5 steps were 110 min each with the following profile: 5 min of 100% buffer A, 2 min of x% buffer C, 3 min of 100% buffer A, a 10 min gradient from 0 to 10% buffer B, and a 90 min gradient from 10 to 50% buffer B. The 2 min buffer C percentages (x) in steps 2-13 were as follows: 10, 30, 50, 70 and 100%. Step 7 is 5 min of 100% buffer A, 2 min of 100% buffer D, 3 min of 100% buffer A, a 10 min gradient from 0 to 10% buffer B, and a 90 min gradient from 10 to 100% buffer B, and a 10 min hold at 100% buffer B.

[0148] The mass spectrometer was operated in a four step cycle, where the 3 most intense ions were scanned in a MS/MS mode (3 μscans per scan). The scan range for the MS experiment was set to m/z 400-2000.

Example 11. Analysis of SEQUEST data

[0149] A singly charged peptide must be tryptic and the cross-correlation score has to be higher than 1.9. Tryptic or partially tryptic peptides with a charge state +2 must have a cross-correlation score of at least 2.2. Peptides with cross-correlation scores (XCorr) above 3 were accepted regardless of their tryptic nature. Triply charged tryptic or partially tryptic peptides were accepted if their XCorr was above 3.75. If proteins were identified by less than 4 different peptide spectra, the existence of the protein was manually checked by at least one good spectrum. Proteins identified by more than 4 peptides were considered as valid identification. Spectra of good quality need to meet the following criteria. MS/MS spectra have to show fragment ions clearly above the noise level with continuity in the b and y ion series. Y-ions of a protein sequence should be intense. The highest and second best scoring amino acid sequence should differ in their cross-correlation score by 0.1 or more.

Results:

[0150] The following data were generated from the application of affinity peptide encoded tags (APEPTags) method on a mixture of six model proteins.

[0151] Qualitative analysis: 35 modified cysteine containing peptides were extracted.

[0152] In the following sequence, "C#" indicates a modified cysteine, and "M@" indicates an oxidized methionine.

```
ALBU_BOVIN - 35 69293
     1  K.CC#TESLVNR.R
     2  K.DAIPENLPPLTADFAEDKDVC#K.N
     3  K.EYEATLEECC#AK.D
     4  K.EYEATLEECC#AKDDPHACYSTVFDK.L
     5  K.LFTFHADIC#TLPDTEK.Q
     6  K.LKEC#CDKPLLEK.S
     7  K.LKPDPNTLC#DEFK.A
     8  R.M@PC#TEDYLSLILNR.L
     9  R.MPC#TEDYLSLILNR.L
    10  R.NEC#FLSHKDDSPDLPK.L
    11  R.RPC#FSALTPDETYVPK.A
    12  K.SHC#IAEVEK.D
    13  K.SLHTLFGDELC#K.V
    14  K.YIC#DNQDTISSK.L
    15  K.YNGVFQECC#QAEDK.G

BGAL_ECOLI -
     1  R.AVVELHTADGTLIEAEAC#DVGFR.E
     2  R.IGLNC#QLAQVAER.V
     3  D.PSRPVQYEGGGADTTATDIIC#PM@YAR.V
     4  D.PSRPVQYEGGGADTTATDIIC#PMYAR.V
     5  R.PVQYEGGGADTTATDIIC#PMYAR.V
     6  K.SVDPSRPVQYEGGGADTTATDIIC#PM@YAR.V
     7  K.SVDPSRPVQYEGGGADTTATDIIC#PMYAR.V

G3P_RABIT -
     1  K.IVSNASC#TTNCLAPLAK.V
     2  K.IVSNASCTTNC#LAPLAK.V
     3  R.VPTPNVSVVDLTC#R.L

LACB_BOVIN -
     1  R.LSFNPTQLEEQC#HI.-

LCA_BOVIN -
     1  K.DDQNPHSSNIC#NISCDK.F
     2  K.DDQNPHSSNICNISC#DK.F
     3  K.FLDDDLTDDIM@C#VK.K
     4  K.FLDDDLTDDIMC#VK.K
     5  K.LDQWLC#EK.L
     6  S.NICNISCDKFLDDDLTDDIMC#VK.K
     7  H.SSNIC#NISCDK.F

OVAL_CHICK - 3 42750
     1  R.ADHPFLFC#IK.H
     2  R.YPILPEYLQC#VK.E
```

[0153] The following data were generated from immobilized peptide encoded tags method, applied to a whole cell extract from yeast. 142 unique proteins were identified.

[0189] Yeast protein extracts:

YAL003W EFB1 1 22627 0.00
1 N.C#VVEDDKVSLDDLQQSIEEDEDHVQSTDIAAMQK.L

YAL005C SSA1 9 69767 0.00
1 K.AVGIDLGTTYSC#VAH.F
2 K.AVGIDLGTTYSC#VAHFANDR.V
3 R.FEELC#ADLFR.S

YAL038W CDC19 20 54545 0.00
1 R.AEVSDVGNAILDGADC#VMLSGETAK.G
2 V.GNAILDGADC#VMLSGETAK.G
3 R.NC#TPKPTSTTETVAASAVAAVFEQK.A
4 K.PVIC#ATQMLESMTYNPR.P
5 K.SNLAGKPVIC#ATQMLESM@TYNPR.P
6 K.SNLAGKPVIC#ATQMLESMTYNPR.P
7 K.YRPNC#PIILVTR.C

YBL024W - 1 77879 0.00
1 R.LVYSTC#SLNPIENEAVVAEALR.K

YBL047C - 1 150783 0.00
1 R.LPNQTLGEIWALC#DR.D

YBL072C RPS8A 2 22490 0.00
1 R.C#DGYILEGEELAFYLR.R

YBL075C SSA3 2 70547 0.00
1 R.AVGIDLGTTYSC#VAHFSNDR.V

YBL087C RPL23A 4 14473 0.00
1 R.ISLGLPVGAIM@NC#ADNSGAR.N
2 R.ISLGLPVGAIMNC#ADNSGAR.N
3 L.PVGAIMNC#ADNSGAR.N

YBR025C - 4 44174 0.00
1 R.C#PLGNPANYPFATIDPEEAR.V
2 K.LDLISFFTC#GPDEVR.E
3 K.PC#IYLINLSER.D
4 R.SVDSIYQVVR.C

YBR031W RPL4A 5 39092 0.00
1 R.SGQGAFGNMC#R.G

YBR048W RPS11B 4 17749 0.00
1 K.C#PFTGLVSIR.G
3 R.VQVGDIVTVGQC#R.P
4 R.VQVGDIVTVGQC#RPISK.T

YBR118W TEF2 17 50033 0.00
1 N.ATVIVLNHPGQISAGYSPVLDC#HTAH.I
2 M.C#VEAFSEYPPLGR.F
3 F.NATVIVLNHPGQISAGYSPVLDC#HTAH.I
4 K.NM@ITGTSQADC#AILIIAGGVGEFEAGISK.
5 K.NMITGTSQADC#AILIIAGGVGEFEAGISK.D
6 K.PMC#VEAFSEYPPLGR.F
7 V.PSKPMC#VEAFSEYPPLGR.F

YBR127C VMA2 4 57749 0.00
1 K.IPIFSASGLPHNEIAAQIC#R.Q

YBR169C SSE2 1 77621 0.00
1 K.GAAFIC#AIHSPTLR.V

YBR249C ARO4 6 39749 0.00
1 K.GNEHC#FVILR.G
2 K.NGTDGTLNVAVDAC#QAAAHSHHFM@GVTK.H
3 K.NGTDGTLNVAVDAC#QAAAHSHHFMGVTK.H
4 R.VLVIVGPC#SIHDLEAAQEYALR.L
5 K.VNDVVC#EQIANGENAITGVMIESNINEGNQGIPAEGK.A
6 K.YGVSITDAC#IGWETTEDVLR.K

YBR263W SHM1 1 62862 0.00
1 K.EISQGC#GAYLMSDMAH.I

YCL009C ILV6 1 33987 0.00
1 K.LVEPFGVLEC#AR.S

YCL030C HIS4 1 87790 0.00
1 K.FHAAQLPTETLEVETQPGVLC#SR.F

YDL014W NOP1 2 34465 0.00
1 R.DHC#IVVGR.Y
2 R.MLIGMVDC#VFADVAQPDQAR.I

YDL055C PSA1 3 39566 0.00
1 K.DNSPFFVLNSDVIC#EYPFK.E
2 K.STIVGWNSTVGQWC#R.L
3 R.SVVLC#NSTIK.N

YDL061C RPS29B 1 6728 0.00
1 R.VC#SSHTGLVR.K

YDL066W IDP1 1 48190 0.00
1 K.C#ATITPDEAR.V

YDL097C RPN6 1 49774 0.00
1 R.SHFNALYDTLLESNLC#K.I

YDL126C CDC48 2 91996 0.00
1 K.DTVLIVLIDDELEDGAC#R.I
2 R.LGDLVTIHPC#PDIK.Y

YDL131W LYS21 3 48594 0.00
1 R.DIENLVADAVEVNIPFNNPITGFC#AF.T
2 R.VGIADTVGC#ANPR.Q

YDL136W RPL35B 1 13910 0.00
1 K.SIAC#VLTVINEQQR.E

YDL229W SSB1 10 66602 0.00
1 G.ERVNC#KENTLLGEFDLKNIPMMPAGEP.V
2 R.TFTTC#ADNQTTVQFPVYQGER.V

YDR002W - 1 22953 0.00
1 K.IC#ANHIIAPEYTLKPNVGSDR.S

YDR035W ARO3 2 41070 0.00
1 R.IMIDC#SHGNSNK.D
2 K.LPIAGEMLDTISPQFLSDC#FSLGAIGAR.T

YDR037W KRS1 2 67959 0.00
1 K.LEC#PPPLTNAR.M

YDR061W - 1 61191 0.00
1 K.YDSIEVSGGC#PIVIGLR.Y

YDR091C - 1 68340 0.00
1 R.APESLLTGC#NR.F

YDR127W ARO1 1 174755 0.00
1 R.ALILAALGEGQC#K.I

YDR155C CPH1 5 17391 0.00
1 N.AGPNTNGSQFFITTVPC#PWLDGK.H
2 M.ANAGPNTNGSQFFITTVPC#PWLDGK.
3 R.PGLLSM@ANAGPNTNGSQFFITTVPC#PWLDGK.H

YDR158W HOM2 1 39544 0.00
1 R.VAVSDGHTEC#ISLR.F

YDR188W CCT6 2 59924 0.00
1 R.AAAAQDEITGDGTTTVVC#LVGELLR.Q
2 R.NAITGATGIASNLLLC#DELLR.A

YDR190C - 2 50453 0.00
1 K.VPFC#PLVGSELYSVEVK.K
2 R.YALQLLAPC#GILAQTSNR.K

YDR226W ADK1 1 24255 0.00
1 K.DELTNNPAC#K.N

YDR321W ASP1 1 41395 0.00
1 K.SQNAAVNGSGIAC#QQR.S

YDR353W TRR1 1 34238 0.00
1 R.NKPLAVIGGGDSAC#EEAQFLTK.Y

YDR385W EFT2 10 93289 0.00
1 R.AEQLYEGPADDANC#IAIK.N
2 K.IWC#FGPDGNGPNLVIDQTK.A
3 R.VTDGALVVVDTIEGVC#VQTETVLR.Q

YDR418W RPL12B 1 17823 0.00
1 K.EILGTAQSVGC#R.V

YDR447C RPS17B 4 15803 0.00
1 R.LC#DEIATIQSK.R

YDR487C RIB3 1 22568 0.00
1 R.GHTEAGVDLC#K.L

YDR502C SAM2 2 42256 0.00
1 K.SLVAAGLC#K.R
2 K.TC#NVLVAIEQQSPDIAQGLHYEK.S

YEL046C GLY1 1 42815 0.00
1 R.THLMQPPYSILC#DYR.A

YEL047C - 2 50844 0.00
1 R.LGGSSLLEC#VVFGR.T

YER007C-A - 1 20278 0.00
1 K.FVLSGANIMC#PGLTSAGADLPPAPGYEK.G
1 K.HYSKPDGPNNNVAVVC#SAR.S

YER055C HIS1 1 32266 0.00
1 K.C#DLGITGVDQVR.E

YER091C MET6 2 85860 0.00
1 K.GMLTGPITC#LR.W

YER107C GLE2 1 40523 0.00
1 R.AQHESSSPVLC#TR.W

YER133W GLC7 2 35907 0.00
1 K.IC#GDIHGQYYDLLR.L
2 K.IFC#MHGGLSPDLNSMEQIR.R

YER177W RPL23B 2 30091 0.00
1 K.SEHQVELIC#SYR.S

YFL018C LPD1 1 54010 0.00
1 K.AAQLGFNTAC#VEK.R

YFL039C ACT1 4 41690 0.00
1 K.LC#YVALDFEQEMQTAAQSSSIEK.S

YFL045C SEC53 4 29063 0.00
1 K.TYC#LQHVEK.D

YGL009C LEU1 4 85794 0.00
1 R.EAEILVVTGDNFGC#GSSR.E
2 K.HC#LVNGLDDIGITLQK.E
3 R.VDC#TLATVDHNIPTESR.K
4 K.VFIGSC#TNGR.I

YGL026C TRP5 3 76626 0.00
1 R.FGDFGGQYVPEALHAC#LR.E
2 K.LPDAVVAC#VGGGSNSTGMFSPFEHDTSVK.L
3 R.LTEHC#QGAQIWLK.R

YGL087C MMS2 1 15545 0.00
1 K.INLPC#VNPTTGEVQTDFHTLR.D

YGL105W ARC1 1 42084 0.00
1 K.STAMVLC#GSNDDKVEFVEPPKDSK.A

YGL135W RPL1B 2 24486 0.00
1 K.SC#GVDAMSVDDLK.K
2 K.SC#GVDAMSVDDLKK.L

YGL147C RPL9A 4 21569 0.00
1 K.DEIVLSGNSVEDVSQNAADLQQIC#R.V
2 N.VKDEIVLSGNSVEDVSQNAADLQQIC#R.V

YGL148W ARO2 3 40838 0.00
1 R.C#PDASVAGLMVK.E
2 K.DSIGGVVTC#VVR.N

YGL157W - 1 38083 0.00
1 K.DC#IVDTAAQMLEVQNEA.-

YGL202W ARO8 1 56178 0.00
1 K.DYFPWDNLSVDSPKPPFPQGIGAPIDEQNC#IK.Y
1 K.C#VHFQNSYYR.K

YGL245W - 1 82663 0.00
1 K.YSAADVAC#WGALR.S

YGR192C TDH3 19 35747 0.00
1 K.IVSNASCTTNC#LAPLAK.V

YGR204W ADE3 2 102205 0.00
1 K.NGHPFFLPC#TPK.G
2 R.SPVTVEDVGC#TGALTALLR.D

YGR234W YHB1 1 44646 0.00
1 K.C#NPNRPIYWIQSSYDEK.T

YGR240C PFK1 2 107970 0.00
1 R.QAAGNLISQGIDALVVC#GGDGSLTGADLFR.H

YGR254W ENO1 7 46816 0.00
2 K.IGLDC#ASSEFFK.D

YGR285C ZUO1 3 49020 0.00
1 R.AQYDSC#DFVADVPPPK.K

YHR019C DED81 2 62207 0.00
2 K.YGTC#PHGGYGIGTER.I

YHR025W THR1 1 38712 0.00
1 K.C#IAIIPQFELSTADSR.G

YHR030C SLT2 1 55636 0.00
1 R.ITVDEALEHPYLSIWHDPADEPVC#SEK.F

YHR064C PDR13 1 62186 0.00
1 K.C#ANGAPAVEVDGK.V

YHR208W BAT1 3 43596 0.00
1 K.EIGWNNEDIHVPLLPGEQC#GALTK.Q
2 R.IC#LPTFESEELIK.L
3 K.LGANYAPC#ILPQLQAAK.R

YHR216W - 1 56530 0.00
1 L.LGGIGFIHHNC#TPEDQADMVR.R

YIL022W TIM44 1 48854 0.00
1 K.LLAPQDIPVLVVGC#R.A

YIL041W - 1 36670 0.00
1 K.VALNSSEC#LNK.M

YIL094C LYS12 1 40069 0.00
1 K.EQC#QGALFGAVQSPTTK.V

YIR006C PAN1 1 160267 0.00
1 R.SIVTNGSNTVSGANC#R.K

YIR034C LYS1 1 41465 0.00
1 R.GGPFDEIPQADIFINC#IYLSK.P

```
YJL045W - 1 69382 0.00
     1  K.YRNVIAHTLDENEC#APVPPAVR.S

YJL130C URA2 1 245126 0.00
     1  R.GHNIPC#TSTISGR.C

YJL138C TIF2 2 44697 0.00
     1  K.VHAC#IGGTSFVEDAEGLR.D

YJL200C - 2 86583 0.00
     1  K.DLPSSIATNQEVFDFLESC#AK.R

YJR016C ILV3 2 62861 0.00
     1  R.EIIADSFETIMMAQHYDANIAIPSC#DK.N
     2  K.LVSNASNGC#VLDA.-

YJR109C CPA2 1 123915 0.00
     1  R.HLGVIGEC#NVQYALQPDGLDYR.V

YJR148W BAT2 2 41625 0.00
     1  R.IC#LPTFDPEELITLIGK.L
     2  K.LGANYAPC#VLPQLQAASR.G

YKL006W RPL14A 1 15167 0.00
     1  K.WAAAAVC#EK.W

YKL060C FBA1 6 39621 0.00
     1  H.MLDLSEETDEENISTC#VK.Y
     2  R.SIAPAYGIPVVLHSDHC#AK.K
     3  K.VNLDTDC#QYAYLTGIR.D

YKL182W FAS1 2 228691 0.00
     1  R.GYTC#QFVDMVLPNTALK.T
     2  R.TC#ILHGPVAAQFTK.V

YKL216W URA1 2 34801 0.00
     1  K.DAFEHLLC#GASMLQIGTELQK.E
     2  K.IQDSEFNGITELNLSC#PNVPGKPQVAYDFDLTK.E

YLL026W HSP104 1 102035 0.00
     1  R.LPDSALDLVDISC#AGVAVAR.D

YLR027C AAT2 1 47793 0.00
     1  K.LSTVSPVFVC#QSFAK.N
     2  K.NPVILADACC#SR.H

YLR058C SHM2 1 52218 0.00
     1  R.M@EILC#QQR.A

YLR075W RPL10 3 25361 0.00
     1  K.MLSC#AGADR.L

YLR109W - 2 19115 0.00
     1  K.FQYIAISQSDADSESC#K.M

YLR153C ACS2 1 75492 0.00
     1  R.TYLPPVSC#DAEDPLFLLYTSGSTGSPK.G

YLR249W YEF3 13 115945 0.00
     1  R.AIANGQVDGFPTQEEC#R.T
     2  R.FIPSLIQC#IADPTEVPETVHLLGATTF.V
```

```
     3  H.IANQSNLSPSVEPYIVQLVPAIC#TNAGNK.D
     5  R.KEIEEHC#SMLGLDPEIVSHSR.I
     6  K.NTYEYEC#SFLLGENIGMK.S
     8  K.PQITDINFQC#SLSSR.I
    10  K.STLINVLTGELLPTSGEVYTHENC#R.I
    13  K.VTNMEFQYPGTSKPQITDINFQC#SLSSR.I

YLR259C HSP60 3 60752 0.00
     1  K.NVAAGC#NPM@DLR.R
     2  K.NVAAGC#NPMDLR.R

YLR304C ACO1 1 85368 0.00
     1  R.VGLIGSC#TNSSYEDMSR.S

YLR355C ILV5 5 44368 0.00
     1  K.YGMDYMYDAC#STTAR.R

YLR441C RPS1A 3 28743 0.00
     1  R.VVEVC#LADLQGSEDHSFR.K

YLR447C VMA6 1 39791 0.00
     1  R.NITWIAEC#IAQNQR.E

YML007W YAP1 1 72533 0.00
     1  S.EFC#SKMNQVCGTRQCPIPKKPISALDK.E

YML008C ERG6 2 43431 0.00
     1  R.GDLVLDVGC#GVGGPAR.E
     2  K.VYAIEATC#HAPK.L

YML028W TSA1 3 21590 0.00
     1  R.LVEAFQWTDKNGTVLPC#NWTPGAATIKPTVEDSK.E
     2  K.NGTVLPC#NWTPGAATIKPTVEDSK.E

YML085C TUB1 1 49800 0.00
     1  K.IGIC#YEPPTATPNSQLATVDR.A

YML126C HMGS 2 55014 0.00
     1  R.VGLFSYGSGLAASLYSC#K.I

YMR079W SEC14 1 34901 0.00
     1  R.AAGHLVETSC#TIMDLK.G

YMR116C BEL1 4 34805 0.00
     1  Q.C#LATLLGHNDWVSQVR.V
     2  K.GQC#LATLLGHNDWVSQVR.V

YMR120C ADE17 1 65263 0.00
     1  K.YTQSNSVC#YAR.N

YMR173W-A - 1 43890 0.00
     1  K.C#PHLEIVNLSDNAFGLR.T

YMR260C TIF11 1 17435 0.00
     1  R.VEASC#FDGNKR.M

YMR315W - 1 38216 0.00
     1  K.IAESTPLPVGVAENWLYLPC#IK.I

YNL104C LEU4 1 68409 0.00
     1  R.GC#GVAATELGMLAGADR.V
```

YNL134C - 1 41164 0.00
1 K.IGPQGALLGC#DAAGQIVK.L

YNL178W RPS3 3 26503 0.00
2 K.GC#EVVVSGK.L

YNL220W ADE12 2 48279 0.00
1 R.C#AGGNNAGHTIVVDGVK.Y
2 R.C#GWLDLVVLK.Y

YNL244C SUI1 1 12312 0.00
1 K.VC#EFMISQLGLQK.K

YNL301C RPL18B 6 20563 0.00
1 K.AGGEC#ITLDQLAVR.A

YNR050C LYS9 6 48918 0.00
1 Y.C#GGLPAPEDSDNPLGYK.F
2 R.GNALDTLC#AR.L
3 F.LSYC#GGLPAPEDSDNPLGYK.F
4 K.SFLSYC#GGLPAPEDSDNPLGYK.F

YOL086C ADH1 5 36849 0.00
2 Y.ATADAVQAAHIPQGTDLAQVAPILC#AGITVYK.A

YOL143C RIB4 1 18556 0.00
1 K.VDMPVIFGLLTC#MTEEQALAR.A

YOR007C SGT2 1 37218 0.00
1 K.EISEDGADSLNVAMDC#ISEAFGFER.E

YOR122C PFY1 1
1 R.HDAEGVVC#VR.T

YOR187W - 1
1 R.ELLNEYGFDGDNAPIIMGSALC#ALEGR.Q

YOR204W DED1 2
1 R.DLMAC#AQTGSGK.T

YOR229W WTM2 1
1 R.FFNNHLFASC#SDDNILR.F

YOR261C RPN8 1
1 R.C#VGVILGDANSSTIR.V

YPL028W ERG10 2
1 K.VNVYGGAVALGHPLGC#SGAR.V

YPL061W ALD6 6
1 K.IAPALAMGNVC#ILK.P
2 K.PAAVTPLNALYFASLC#K.K

YPL117C IDI1 1
1 K.IIC#ENYLFNWWEQLDDLSEVENDR.Q

Totals: # Unique Proteins = 142
Unique Peptides = 218

```
YNL178W RPS3 3 26503 0.00
     2  K.GC#EVVVSGK.L

YNL220W ADE12 2 48279 0.00
     1  R.C#AGGNNAGHTIVVDGVK.Y
     2  R.C#GWLDLVVLK.Y

YNL244C SUI1 1 12312 0.00
     1  K.VC#EFMISQLGLQK.K

YNL301C RPL18B 6 20563 0.00
     1  K.AGGEC#ITLDQLAVR.A

YNR050C LYS9 6 48918 0.00
     1  Y.C#GGLPAPEDSDNPLGYK.F
     2  R.GNALDTLC#AR.L
     3  F.LSYC#GGLPAPEDSDNPLGYK.F
     4  K.SFLSYC#GGLPAPEDSDNPLGYK.F

YOL086C ADH1 5 36849 0.00
     2  Y.ATADAVQAAHIPQGTDLAQVAPILC#AGITVYK.A

YOL143C RIB4 1 18556 0.00
     1  K.VDMPVIFGLLTC#MTEEQALAR.A

YOR007C SGT2 1 37218 0.00
     1  K.EISEDGADSLNVAMDC#ISEAFGFER.E

YOR122C PFY1 1
     1  R.HDAEGVVC#VR.T

YOR187W - 1
     1  R.ELLNEYGFDGDNAPIIMGSALC#ALEGR.Q

YOR204W DED1 2
     1  R.DLMAC#AQTGSGK.T

YOR229W WTM2 1
     1  R.FFNNHLFASC#SDDNILR.F

YOR261C RPN8 1
     1  R.C#VGVILGDANSSTIR.V

YPL028W ERG10 2
     1  K.VNVYGGAVALGHPLGC#SGAR.V

YPL061W ALD6 6
     1  K.IAPALAMGNVC#ILK.P
     2  K.PAAVTPLNALYFASLC#K.K

YPL117C IDI1 1
     1  K.IIC#ENYLFNWWEQLDDLSEVENDR.Q

Totals:  # Unique Proteins = 142
         # Unique Peptides = 218
```

CONCLUSION

[0154] Thus, it will be appreciated that the compounds and methods described herein are used to identify proteins using mass spectrometry.

[0155] One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The molecular complexes and the methods, procedures, molecules, and specific compounds described herein are presently representative of preferred embodiments and are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention and are defined by the scope of the claims.

[0156] It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

[0157] All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains. All patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

[0158] The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising", "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that the use of such terms and expressions indicates the exclusion of equivalents of the features shown and described or portions thereof. It is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

[0159] In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. For example, if X is described as selected from the group consisting of bromine, chlorine, and iodine, claims for X being bromine and claims for X being bromine and chlorine are fully described.